Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 7 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review (e20708)
Julia Höhn, Achim Hekler, Eva Krieghoff-Henning, Jakob Kather, Jochen Utikal, Friedegund Meier, Frank Gellrich, Axel Hauschild, Lars French, Justin Schlager, Kamran Ghoreschi, Tabea Wilhelm, Heinz Kutzner, Markus Heppt, Sebastian Haferkamp, Wiebke Sondermann, Dirk Schadendorf, Bastian Schilling, Roman Maron, Max Schmitt, Tanja Jutzi, Stefan Fröhling, Daniel Lipka, Titus Brinker. 11

Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis (e21680)
Ardalan Mirzaei, Parisa Aslani, Edward Luca, Carl Schneider. 25

Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis (e23863)
Jo-Hsuan Wu, T Liu, Wan-Ting Hsu, Jennifer Ho, Chien-Chang Lee. 39

Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-analysis (e27370)
Scarlet Nazarian, Ben Glover, Hutan Ashrafian, Ara Darzi, Julian Teare. 54

Implementation of Telemental Health Services Before COVID-19: Rapid Umbrella Review of Systematic Reviews (e26492)
Phoebe Barnett, Lucy Goulding, Cecilia Casetta, Harriet Jordan, Luke Sheridan-Rains, Thomas Steare, Julie Williams, Lisa Wood, Fiona Gaughran, Sonia Johnson. 72

Exploring the Role of Web-Based Interventions in the Self-management of Dementia: Systematic Review and Narrative Synthesis (e26551)
Abigail Lee, Esther Gerritzen, Orii McDermott, Martin Orrell. 117

Effectiveness of eHealth Interventions in Improving Medication Adherence for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review (e29475)
Mieke Schulte, Jiska Aardoom, Lisa Loheide-Niesmann, Leonie Verstraete, Hans Ossebaard, Heleen Riper. 129

The Use of Twitter by Medical Journals: Systematic Review of the Literature (e26378)
Natalie Erskine, Sharief Hendricks. 142

Social Media and mHealth Technology for Cancer Screening: Systematic Review and Meta-analysis (e26759)
Arlinda Ruco, Fahima Dossa, Jill Tinmouth, Diego Llovet, Jenna Jacobson, Teruko Kishibe, Nancy Baxter. 152

XSL•FO
RenderX

The Efficacy of eHealth Interventions for the Treatment of Adults Diagnosed With Full or Subthreshold Binge Eating Disorder: Systematic Review and Meta-analysis (e17874)
Elnaz Moghimi, Caroline Davis, Michael Rotondi. . . . 252

Functionalities and Issues in the Implementation of Personal Health Records: Systematic Review (e26236)
Nabila Harahap, Putu Handayani, Achmad Hidayanto. . . . 765

Recommendations for the Design and Implementation of Virtual Reality for Acquired Brain Injury Rehabilitation: Systematic Review (e26344)
Sophie Brassel, Emma Power, Andrew Campbell, Melissa Brunner, Leanne Togher. . . . 1100

Characteristics of the Measurement Tools for Assessing Health Information–Seeking Behaviors in Nationally Representative Surveys: Systematic Review (e27539)
Hanna Choi, Gyeonghui Jeong. . . . 1135

Viewpoints

The Ethical Digital Surgeon (e25849)
Kyle Lam, Sanjay Purkayastha, James Kinross. . . . 178

Overcoming the Digital Divide in the Post–COVID-19 “Reset”: Enhancing Group Virtual Visits with Community Health Workers (e27682)
Megha Shah, Ashley Gibbs, Mohammed Ali, K Narayan, Nadia Islam. . . . 181

Original Papers

Time to Change for Mental Health and Well-being via Virtual Professional Coaching: Longitudinal Observational Study (e27774)
Alexis Jeannotte, Derek Hutchinson, Gabriella Kellerman. . . . 187

Efficacy of a Just-in-Time Adaptive Intervention to Promote HIV Risk Reduction Behaviors Among Young Adults Experiencing Homelessness: Pilot Randomized Controlled Trial (e26704)
Diane Santa Maria, Nikhil Padhye, Michael Businelle, Yijiong Yang, Jennifer Jones, Alexis Sims, Marguerita Lightfoot. . . . 203

Behavioral Activation and Depression Symptomatology: Longitudinal Assessment of Linguistic Indicators in Text-Based Therapy Sessions (e28244)
Hannah Burkhardt, George Alexopoulos, Michael Pullmann, Thomas Hull, Patricia Areán, Trevor Cohen. . . . 221

A Digital Intervention for Primary Care Practitioners to Support Antidepressant Discontinuation (Advisor for Health Professionals): Development Study (e25537)
Hannah Bowers, Tony Kendrick, Nadja van Ginneken, Marta Glowacka, Samantha Williams, Geraldine Leydon, Carl May, Christopher Dowrick, Joanna Moncrieff, Chris Johnson, Michael Moore, Rebecca Laine, Adam Geraghty. . . . 237

Efficacy, Use, and Acceptability of a Web-Based Self-management Intervention Designed to Maximize Sexual Well-being in Men Living With Prostate Cancer: Single-Arm Experimental Study (e21502)
Sean O'Connor, Carrie Flannagan, Kader Parahoo, Mary Steele, Samantha Thompson, Suneil Jain, Michael Kirby, Nuala Brady, Roma Maguire, John Connaghan, Eilis McCaughan. . . . 263

XSL•FO
RenderX

Older Adults' Experiences of Behavior Change Support in a Digital Fall Prevention Exercise Program: Qualitative Study Framed by the Self-determination Theory (e26235)
Beatrice Pettersson, Rebecka Janols, Maria Wiklund, Lillemor Lundin-Olsson, Marlene Sandlund. . . . 280

Supporting Patient-Clinician Interaction in Chronic HIV Care: Design and Development of a Patient-Reported Outcomes Software Application (e27861)
Susan Herrmann, Brad Power, Amineh Rashidi, Mark Cypher, Frank Mastaglia, Amy Grace, Elizabeth McKinnon, Pierre Sarrot, Christophe Michau, Matthew Skinner, Renae Desai, Martin Duracinsky. . . . 293

Approaches for Implementing App-Based Digital Treatments for Drug Use Disorders Into Primary Care: A Qualitative, User-Centered Design Study of Patient Perspectives (e25866)
Joseph Glass, Theresa Matson, Catherine Lim, Andrea Hartzler, Kilian Kimbel, Amy Lee, Tara Beatty, Rebecca Parrish, Ryan Caldeiro, Angela Garza McWethy, Geoffrey Curran, Katharine Bradley. . . . 306

An Unstructured Supplementary Service Data System for Daily Tracking of Patient Samples and Diagnostic Results in a Diagnostic Network in Malawi: System Development and Field Trial (e26582)
Daniel Killian, Emma Gibson, Mphatso Kachule, Kara Palamountain, Joseph Bangoh, Sarang Deo, Jonas Jonasson. . . . 320

App-Based Feedback for Rehabilitation Exercise Correction in Patients With Knee or Hip Osteoarthritis: Prospective Cohort Study (e26658)
Johanna Biebl, Marzena Rykala, Maximilian Strobel, Pawandeep Kaur Bollinger, Bernhard Ulm, Eduard Kraft, Stephan Huber, Andreas Lorenz.
3 3 2

Acceptability of an mHealth App That Provides Harm Reduction Services Among People Who Inject Drugs: Survey Study (e25428)
Tyler Shelby, Xin Zhou, Douglas Barber, Frederick Altice. . . . 345

Use of Ecological Momentary Assessment Through a Passive Smartphone-Based App (eB2) by Patients With Schizophrenia: Acceptability Study (e26548)
Javier-David Lopez-Morinigo, María Barrigón, Alejandro Porras-Segovia, Verónica Ruiz-Ruano, Adela Escribano Martínez, Paula Escobedo-Aedo, Sergio Sánchez Alonso, Laura Mata Iturralde, Laura Muñoz Lorenzo, Antonio Artés-Rodríguez, Anthony David, Enrique Baca-García. . . . 359

App-Tailoring Requirements to Increase Stress Management Competencies Within Families: Cross-sectional Survey Study (e26376)
Laura Bischoff, Hannes Baumann, Charlotte Meixner, Patricia Nixon, Bettina Wollesen. . . . 372

Self-guided Cognitive Behavioral Therapy Apps for Depression: Systematic Assessment of Features, Functionality, and Congruence With Evidence (e27619)
Laura Martinengo, Anne-Claire Stona, Konstadina Griva, Paola Dazzan, Carmine Pariante, Florian von Wangenheim, Josip Car. . . . 387

A Smart Diaper System Using Bluetooth and Smartphones to Automatically Detect Urination and Volume of Voiding: Prospective Observational Pilot Study in an Acute Care Hospital (e29979)
Jae Cho, Jung-Yeon Choi, Nak-Hyun Kim, Yejee Lim, Jung Ohn, Eun Kim, Jiwon Ryu, Jangsun Kim, Yiseob Kim, Sun-wook Kim, Kwang-Il Kim. . . . 404

Leading Topics in Twitter Discourse on JUUL and Puff Bar Products: Content Analysis (e26510)
Jon-Patrick Allem, Allison Dormanesh, Anuja Majmundar, Vanessa Rivera, Maya Chu, Jennifer Unger, Tess Cruz. . . . 414

Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis (e25925)
Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Nadia Sanz Lamora, Aida Álvarez, Alexandre González-Rodríguez, Meritxell Lozano Rochel, Roger Llobet Vives, Diego Velazquez, Josep Gonfaus, Jordi Gonzàlez. . . . 421

Global Public Interests and Dynamic Trends in Osteoporosis From 2004 to 2019: Infodemiology Study (e25422)
Peng Wang, Qing Xu, Rong-Rong Cao, Fei-Yan Deng, Shu-Feng Lei. . . . 461

XSL•FO
RenderX

Using Search Trends to Analyze Web-Based Interest in Lower Urinary Tract Symptoms-Related Inquiries, Diagnoses, and Treatments in Mainland China: Infodemiology Study of Baidu Index Data (e27029)
Shanzun Wei, Ming Ma, Changjing Wu, Botao Yu, Lisha Jiang, Xi Wen, Fudong Fu, Ming Shi. 470

Prediction of Asthma Hospitalizations for the Common Cold Using Google Trends: Infodemiology Study (e27044)
Bernardo Sousa-Pinto, Jaana Halonen, Aram Antó, Vesa Jormanainen, Wienczyslawa Czarlewski, Anna Bedbrook, Nikolaos Papadopoulos, Alberto Freitas, Tari Haahtela, Josep Antó, João Fonseca, Jean Bousquet. 487

Assessing Patient Perceptions and Experiences of Paracetamol in France: Infodemiology Study Using Social Media Data Mining (e25049)
Stéphane Schück, Avesta Roustamal, Anaïs Gedik, Paméla Voillot, Pierre Foulquié, Catherine Penfornis, Bernard Job. 502

Using Application Programming Interfaces (APIs) to Access Google Data and Gain Insights Into Searches on Birth Control in Louisiana and Mississippi, 2014-2018: Infoveillance Study (e25923)
Ndola Prata, Karen Weidert, Anne Zepecki, Elina Yon, Elizabeth Pleasants, Petrice Sams-Abiodun, Sylvia Guendelman. 516

The Influence of Online Health Information Seeking Before a Consultation on Anxiety, Satisfaction, and Information Recall, Mediated by Patient Participation: Field Study (e23670)
Melanie de Looper, Julia van Weert, Barbara Schouten, Sifra Bolle, Eric Belgers, Eric Eddes, Ellen Smets. 528

How Patient Work Changes Over Time for People With Multimorbid Type 2 Diabetes: Qualitative Study (e25992)
Kathleen Yin, Joshua Jung, Enrico Coiera, Kenneth Ho, Sanjyot Vagholkar, Ann Blandford, Frances Rapport, Annie Lau. 546

Human Enough: A Qualitative Study of Client Experience With Internet-Based Access to Pre-exposure Prophylaxis (e22650)
Shana Hughes, Kimberly Koester, Edvard Engesaeth, Merissa Hawkins, Robert Grant. 559

Using Telerehabilitation to Deliver a Home Exercise Program to Youth With Arthrogryposis: Single Cohort Pilot Study (e27064)
Marianne Gagnon, Gabriela Marino Merlo, Rita Yap, Jessica Collins, Caroline Elfassy, Bonita Sawatzky, Jacquelyn Marsh, Reggie Hamdy, Louis-Nicolas Veilleux, Noémi Dahan-Oliel. 571

Application of Telemedicine Services Based on a Regional Telemedicine Platform in China From 2014 to 2020: Longitudinal Trend Analysis (e28009)
Fangfang Cui, Xianying He, Yunkai Zhai, Minzhao Lyu, Jinming Shi, Dongxu Sun, Shuai Jiang, Chenchen Li, Jie Zhao. 584

Clinical Outcomes of Asynchronous Versus Synchronous Telepsychiatry in Primary Care: Randomized Controlled Trial (e24047)
Peter Yellowlees, Michelle Parish, Alvaro Gonzalez, Steven Chan, Donald Hilty, Byung-Kwang Yoo, J Leigh, Robert McCarron, Lorin Scher, Andres Sciolla, Jay Shore, Glen Xiong, Katherine Soltero, Alice Fisher, Jeffrey Fine, Jennifer Bannister, Ana-Maria Iosif. 597

Using a Real-Time Locating System to Evaluate the Impact of Telemedicine in an Emergency Department During COVID-19: Observational Study (e29240)
Birju Patel, Stacie Vilendrer, Samantha Kling, Ian Brown, Ryan Ribeira, Matthew Eisenberg, Christopher Sharp. 612

Effectiveness of a Cloud-Based Telepathology System in China: Large-Sample Observational Study (e23799)
Xianying He, Linlin Wang, Li Wang, Jinghong Gao, Fangfang Cui, Qianqian Ma, Wenjie Zhang, Lin Wang, Yunkai Zhai, Jie Zhao. 621

Digital Technology Use and BMI: Evidence From a Cross-sectional Analysis of an Adolescent Cohort Study (e26485)
Chen Shen, Iroise Dumontheil, Michael Thomas, Martin Röösli, Paul Elliott, Mireille Toledano. 634

XSL•FO
RenderX

Web-Based Health Information Seeking Among African American and Hispanic Men Living With Chronic Conditions: Cross-sectional Survey Study (e26180)
Ledric Sherman, Kirby Goidel, Caroline Bergeron, Matthew Smith. . . . 650

Perceptions Toward the Use of Digital Technology for Enhancing Family Planning Services: Focus Group Discussion With Beneficiaries and Key Informative Interview With Midwives (e25947)
Hind Yousef, Nihaya Al-Sheyab, Mohannad Al Nsour, Yousef Khader, Malika Al Kattan, Marco Bardus, Mohammad Alyahya, Hana Taha, Mirwais Amiri. . . . 666

Measurement of Cancer-Related Fatigue Based on Heart Rate Variability: Observational Study (e25791)
Chi-Huang Shih, Pai-Chien Chou, Ting-Ling Chou, Tsai-Wei Huang. . . . 677

Experience Sampling and Programmed Intervention Method and System for Planning, Authoring, and Deploying Mobile Health Interventions: Design and Case Reports (e24278)
Bruna Cunha, Kamila Rodrigues, Isabela Zaine, Elias da Silva, Caio Viel, Maria Pimentel. . . . 689

Using Text Mining Techniques to Identify Health Care Providers With Patient Safety Problems: Exploratory Study (e19064)
Iris Hendrickx, Tim Voets, Pieter van Dyk, Rudolf Kool. . . . 717

The Challenges and Pitfalls of Detecting Sleep Hypopnea Using a Wearable Optical Sensor: Comparative Study (e24171)
Zhongxing Zhang, Ming Qi, Gordana Hügli, Ramin Khatami. . . . 726

Empowering Patients Living With Chronic Conditions Using Video as an Educational Tool: Scoping Review (e26427)
Olga Navarro, Marta Escrivá, Raquel Faubel, Vicente Traver. . . . 742

Pharmacy Customers' Experiences of Use, Usability, and Satisfaction of a Nationwide Patient Portal: Survey Study (e25368)
Maria Sääskilahti, Riitta Ahonen, Johanna Timonen. . . . 751

Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)–Based Data Model: Development and Topic Modeling Study (e26770)
Amrita De, Ming Huang, Tinghao Feng, Xiaomeng Yue, Lixia Yao. . . . 789

Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation (e27633)
András Lánczky, Balázs Gy rffy. . . . 803

Changes in Language Style and Topics in an Online Eating Disorder Community at the Beginning of the COVID-19 Pandemic: Observational Study (e28346)
Johannes Feldhege, Markus Moessner, Markus Wolf, Stephanie Bauer. . . . 810

A Newly Developed Online Peer Support Community for Depression (Depression Connect): Qualitative Study (e25917)
Dorien Smit, Janna Vrijsen, Bart Groeneweg, Amber Vellinga-Dings, Janneke Peelen, Jan Spijker. . . . 822

Support Seeking in the Postpartum Period: Content Analysis of Posts in Web-Based Parenting Discussion Groups (e26600)
Bonnie Chivers, Rhonda Garad, Lisa Moran, Siew Lim, Cheryce Harrison. . . . 840

Developing a Time-Adaptive Prediction Model for Out-of-Hospital Cardiac Arrest: Nationwide Cohort Study in Korea (e28361)
Ji Kim, Juhyung Ha, Taerim Kim, Hee Yoon, Sung Hwang, Ik Jo, Tae Shin, Min Sim, Kyunga Kim, Won Cha. . . . 851

XSL•FO
RenderX

Effective Treatment Recommendations for Type 2 Diabetes Management Using Reinforcement Learning: Treatment Recommendation Model Development and Validation (e27858)
Xingzhi Sun, Yong Bee, Shao Lam, Zhuo Liu, Wei Zhao, Sing Chia, Hanis Abdul Kadir, Jun Wu, Boon Ang, Nan Liu, Zuo Lei, Zhuoyang Xu, Tingting Zhao, Gang Hu, Guotong Xie. . . . 862

Frustration With Technology and its Relation to Emotional Exhaustion Among Health Care Workers: Cross-sectional Observational Study (e26817)
Daniel Tawfik, Amrita Sinha, Mohsen Bayati, Kathryn Adair, Tait Shanafelt, J Sexton, Jochen Profit. . . . 879

Efficacy and Cost-effectiveness of Promotion Methods to Recruit Participants to an Online Screening Registry for Alzheimer Disease Prevention Trials: Observational Study (e26284)
Kenichiro Sato, Yoshiki Niimi, Ryoko Ihara, Kazushi Suzuki, Tatsushi Toda, Atsushi Iwata, Takeshi Iwatsubo. . . . 892

A Health Information Quality Assessment Tool for Korean Online Newspaper Articles: Development Study (e24436)
Naae Lee, Seung-Won Oh, Belong Cho, Seung-Kwon Myung, Seung-Sik Hwang, Goo Yoon. . . . 905

Content Analysis and Quality Evaluation of Cesarean Delivery–Related Videos on YouTube: Cross-sectional Study (e24994)
Kyong-No Lee, Yeon Joo, So Choi, Sung Park, Keun-Young Lee, Youngmi Kim, Ga-Hyun Son. . . . 915

US Privacy Laws Go Against Public Preferences and Impede Public Health and Research: Survey Study (e25266)
Cason Schmit, Theodoros Giannouchos, Mahin Ramezani, Qi Zheng, Michael Morrisey, Hye-Chung Kum. . . . 925

Hospital Productivity After Data Breaches: Difference-in-Differences Analysis (e26157)
Jinhyung Lee, Sung Choi. . . . 938

The Jordan Stillbirth and Neonatal Mortality Surveillance (JSANDS) System: Evaluation Study (e29143)
Yousef Khader, Mohammad Alyahya, Ziad El-Khatib, Anwar Batieha, Nihaya Al-Sheyab, Khulood Shattnawi. . . . 946

Recruitment and Retention Strategies Among Racial and Ethnic Minorities in Web-Based Intervention Trials: Retrospective Qualitative Analysis (e23959)
DaSol Hwang, Alex Lee, Jae Song, Hae-Ra Han. . . . 955

Measuring Success of Patients' Continuous Use of Mobile Health Services for Self-management of Chronic Conditions: Model Development and Validation (e26670)
Ting Song, Ning Deng, Tingru Cui, Siyu Qian, Fang Liu, Yingping Guan, Ping Yu. . . . 968

The Effects of Continuous Usage of a Diabetes Management App on Glycemic Control in Real-world Clinical Practice: Retrospective Analysis (e23227)
Yu-Zhen Tu, Ya-Ting Chang, Hung-Yi Chiou, Ken Lai. . . . 982

Uptake of and Engagement With an Online Sexual Health Intervention (HOPE eIntervention) Among African American Young Adults: Mixed Methods Study (e22203)
Alicia Williamson, Andrea Barbarin, Bettina Campbell, Terrance Campbell, Susan Franzen, Thomas Reischl, Marc Zimmerman, Tiffany Veinot.
9 9 2

A Brief Intervention to Increase Uptake and Adherence of an Internet-Based Program for Depression and Anxiety (Enhancing Engagement With Psychosocial Interventions): Randomized Controlled Trial (e23029)
Philip Batterham, Alison Calear, Matthew Sunderland, Frances Kay-Lambkin, Louise Farrer, Helen Christensen, Amelia Gulliver. . . . 1010

XSL•FO
RenderX

Turn on, Tune in, and Drop out: Predictors of Attrition in a Prospective Observational Cohort Study on Psychedelic Use (e25973)
Sebastian Hübner, Eline Haijen, Mendel Kaelen, Robin Carhart-Harris, Hannes Kettner. 1033

Preadolescent Students' Engagement With an mHealth Intervention Fostering Social Comparison for Health Behavior Change: Crossover Experimental Study (e21202)
Raoul Nuijten, Pieter Van Gorp, Tom Borghouts, Pascale Le Blanc, Pauline Van den Berg, Astrid Kemperman, Ehsan Hadian, Monique Simons.
1 0 4 7

Guided Relaxation–Based Virtual Reality for Acute Postoperative Pain and Anxiety in a Pediatric Population: Pilot Observational Study (e26328)
Vanessa Olbrecht, Keith O'Conor, Sara Williams, Chloe Boehmer, Gilbert Marchant, Susan Glynn, Kristie Geisler, Lili Ding, Gang Yang, Christopher King. 1061

The Effects of Immersive Virtual Reality in Reducing Public Stigma of Mental Illness in the University Population of Hong Kong: Randomized Controlled Trial (e23683)
Anna Yuen, Winnie Mak. 1073

Resuscitating Cardiopulmonary Resuscitation Training in a Virtual Reality: Prospective Interventional Study (e22920)
Janaya Perron, Michael Coffey, Andrew Lovell-Simons, Luis Dominguez, Mark King, Chee Ooi. 1090

Unique Internet Search Strategies of Individuals With Self-Stated Autism: Quantitative Analysis of Search Engine Users' Investigative Behaviors (e23829)
Eldad Yechiam, Elad Yom-Tov. 1124

Barriers to Dissemination of Local Health Data Faced by US State Agencies: Survey Study of Behavioral Risk Factor Surveillance System Coordinators (e16750)
Manik Ahuja, Robert Aseltine Jr. 1156

One Decade of Online Patient Feedback: Longitudinal Analysis of Data From a German Physician Rating Website (e24229)
Martin Emmert, Stuart McLennan. 1163

Determinants of Fitness App Usage and Moderating Impacts of Education-, Motivation-, and Gamification-Related App Features on Physical Activity Intentions: Cross-sectional Survey Study (e26063)
Yanxiang Yang, Joerg Koenigstorfer. 1179

The Effect of 24/7, Digital-First, NHS Primary Care on Acute Hospital Spending: Retrospective Observational Analysis (e24917)
Sam Winward, Tejal Patel, Mazin Al-Saffar, Matthew Noble. 1203

The Sharing Economy in China's Aging Industry: Applications, Challenges, and Recommendations (e27758)
Yaolin Hu, Jian Wang, Stephen Nicholas, Elizabeth Maitland. 1215

Recursive Partitioning vs Computerized Adaptive Testing to Reduce the Burden of Health Assessments in Cleft Lip and/or Palate: Comparative Simulation Study (e26412)
Conrad Harrison, Chris Sidey-Gibbons, Anne Klassen, Karen Wong Riff, Dominic Furniss, Marc Swan, Jeremy Rodrigues. 1228

Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study (e26151)
Stanislav Nikolov, Sam Blackwell, Alexei Zverovitch, Ruheena Mendes, Michelle Livne, Jeffrey De Fauw, Yojan Patel, Clemens Meyer, Harry Askham, Bernadino Romera-Paredes, Christopher Kelly, Alan Karthikesalingam, Carlton Chu, Dawn Carnell, Cheng Boon, Derek D'Souza, Syed Moinuddin, Bethany Garie, Yasmin McQuinlan, Sarah Ireland, Kiarna Hampton, Krystle Fuller, Hugh Montgomery, Geraint Rees, Mustafa Suleyman, Trevor Back, Cían Hughes, Joseph Ledsam, Olaf Ronneberger. 1236

XSL•FO
RenderX

Application of an Anomaly Detection Model to Screen for Ocular Diseases Using Color Retinal Fundus Images: Design and Evaluation Study (e27822)
Yong Han, Weiming Li, Mengmeng Liu, Zhiyuan Wu, Feng Zhang, Xiangtong Liu, Lixin Tao, Xia Li, Xiuhua Guo. 1258

Relative Performance of Machine Learning and Linear Regression in Predicting Quality of Life and Academic Performance of School Children in Norway: Data Analysis of a Quasi-Experimental Study (e22021)
Robert Froud, Solveig Hansen, Hans Ruud, Jonathan Foss, Leila Ferguson, Per Fredriksen. 1271

Studying Microtemporal, Within-Person Processes of Diet, Physical Activity, and Related Factors Using the APPetite-Mobile-App: Feasibility, Usability, and Validation Study (e25850)
Alea Ruf, Elena Koch, Ulrich Ebner-Priemer, Monika Knopf, Andreas Reif, Silke Matura. 1286

Australian Children's Exposure to, and Engagement With, Web-Based Marketing of Food and Drink Brands: Cross-sectional Observational Study (e28144)
Bridget Kelly, Rebecca Bosward, Becky Freeman. 1303

Blockchain for Increased Trust in Virtual Health Care: Proof-of-Concept Study (e28496)
Anton Hasselgren, Jens-Andreas Hanssen Rensaa, Katina Kralevska, Danilo Gligoroski, Arild Faxvaag. 1318

Policy Interventions, Development Trends, and Service Innovations of Internet Hospitals in China: Documentary Analysis and Qualitative Interview Study (e22330)
Yunfeng Lai, Shengqi Chen, Meng Li, Carolina Ung, Hao Hu. 1333

Understanding “Internet Plus Healthcare” in China: Policy Text Analysis (e23779)
Feng Yang, Huilin Shu, Xiaoqian Zhang. 1347

Consultation Pricing of the Online Health Care Service in China: Hierarchical Linear Regression Approach (e29170)
Ya-Ling Chiu, Jying-Nan Wang, Haiyan Yu, Yuan-Teng Hsu. 1358

Status Quo and Research Trends of Neurosurgical Departments in China: Bibliometric and Scientometric Analyses (e25700)
Bowen Ni, Minyi He, Bei Cao, Jianmin He, Yawei Liu, Zhen Zhao. 1377

Predictive Modeling of Morbidity and Mortality in Patients Hospitalized With COVID-19 and its Clinical Implications: Algorithm Development and Interpretation (e29514)
Joshua Wang, Wenke Liu, Xiaoshan Chen, Michael McRae, John McDevitt, David Fenyö. 1403

Racial and Ethnic Differences in COVID-19 Outcomes, Stressors, Fear, and Prevention Behaviors Among US Women: Web-Based Cross-sectional Study (e26296)
Jamila Stockman, Brittany Wood, Katherine Anderson. 1419

Privacy-Oriented Technique for COVID-19 Contact Tracing (PROTECT) Using Homomorphic Encryption: Design and Development Study (e26371)
Yongdae An, Seungmyung Lee, Seungwoo Jung, Howard Park, Yongsoo Song, Taehoon Ko. 1427

Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study (e28615)
Colton Margus, Natasha Brown, Attila Hertelendy, Michelle Safferman, Alexander Hart, Gregory Ciottone. 1442

COVID-19 Information Dissemination Using the WeChat Communication Index: Retrospective Analysis Study (e28563)
Zina Fan, Wenqiang Yin, Han Zhang, Dandan Wang, Chengxin Fan, Zhongming Chen, Jinwei Hu, Dongping Ma, Hongwei Guo. 1460

XSL•FO
RenderX

Online Interactive Platform for COVID-19 Literature Visual Analytics: Platform Development Study (e26995)
Addy Moran, Shawn Hampton, Scott Dowson, John Dagdelen, Amalie Trewartha, Gerbrand Ceder, Kristin Persson, Elise Saxon, Andrew Barker, Lauren Charles, Bobbie-Jo Webb-Robertson. 1471

Evaluation of a Social Media Campaign in Saskatchewan to Promote Healthy Eating During the COVID-19 Pandemic: Social Media Analysis and Qualitative Interview Study (e27448)
Jordyn Grantham, Carrie Verishagen, Susan Whiting, Carol Henry, Jessica Lieffers. 1481

Risk Factors of Psychological Responses of Chinese University Students During the COVID-19 Outbreak: Cross-sectional Web-Based Survey Study (e29312)
Xudong Zhang, Xin Shi, Yang Wang, Huiquan Jing, Qingqing Zhai, Kunhang Li, Dan Zhao, Shiyu Zhong, Yuequn Song, Feng Zhang, Yijun Bao. 1495

Age and Social Disparities in the Use of Telemedicine During the COVID-19 Pandemic in Japan: Cross-sectional Study (e27982)
Atsushi Miyawaki, Takahiro Tabuchi, Michael Ong, Yusuke Tsugawa. 1506

Evaluation of a Parsimonious COVID-19 Outbreak Prediction Model: Heuristic Modeling Approach Using Publicly Available Data Sets (e28812)
Agrayan Gupta, Shaun Grannis, Suranga Kasthurirathne. 1520

Public Awareness and Utilization of 937-Telephone Health Services in the Kingdom of Saudi Arabia Before and During the COVID-19 Pandemic: Longitudinal Study (e27618)
Saja Al-Rayes, Arwa Alumran, Duaa Aljabri, Afnan Aljaffary, Ethar Aldoukhi, Zainab Alahmedalyousif, Reem Al Madani. 1529

Online-Delivered Group and Personal Exercise Programs to Support Low Active Older Adults' Mental Health During the COVID-19 Pandemic: Randomized Controlled Trial (e30709)
Mark Beauchamp, Ryan Hulteen, Geralyn Ruissen, Yan Liu, Ryan Rhodes, Colin Wierts, Katrina Waldhauser, Samantha Harden, Eli Puterman.
1 5 3 9

Stakeholder Perspectives on Barriers and Facilitators for the Adoption of Virtual Clinical Trials: Qualitative Study (e26813)
Romée Coert, James Timmis, André Boorsma, Wilrike Pasman. 1553

Corrigenda and Addendas

Correction: Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment (e31553)
Armando Rotondi, Jonathan Grady, Barbara Hanusa, Gretchen Haas, Michael Spring, Kaleab Abebe, James Luther, John Gurklis. 1389

Correction: A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial (e31543)
Richard Morriss, Catherine Kaylor-Hughes, Matthew Rawsthorne, Neil Coulson, Sandra Simpson, Boliang Guo, Marilyn James, James Lathe, Paul Moran, Laila Tata, Laura Williams. 1391

Correction: Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource (e31544)
Konrad Krawczyk, Tadeusz Chelkowski, Daniel Laydon, Swapnil Mishra, Denise Xifara, Benjamin Gibert, Seth Flaxman, Thomas Mellan, Veit Schwämmle, Richard Röttger, Johannes Hadsund, Samir Bhatt. 1393

Correction: Olfactory Training and Visual Stimulation Assisted by a Web Application for Patients With Persistent Olfactory Dysfunction After SARS-CoV-2 Infection: Observational Study (e32120)
Fabrice Denis, Anne-Lise Septans, Lea Periers, Jean-Michel Maillard, Florian Legoff, Hirac Gurden, Sylvain Moriniere. 1395

XSL•FO
RenderX

Correction: A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study (e31085)
Cheng-Sheng Yu, Shy-Shin Chang, Tzu-Hao Chang, Jenny Wu, Yu-Jiun Lin, Hsiung-Fei Chien, Ray-Jade Chen. 1397

Correction: Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study (e31481)
Lizong Deng, Luming Chen, Tao Yang, Mi Liu, Shicheng Li, Taijiao Jiang. 1399

Correction: A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology (e31601)
Ji-Hyun Lee, Hyeoun-Ae Park, Tae-Min Song. 1401

XSL•FO
RenderX

Review

Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review

Julia Höhn[1*], MSc; Achim Hekler[1*], MSc; Eva Krieghoff-Henning[1], PhD; Jakob Nikolas Kather[2,3], Prof Dr med; Jochen Sven Utikal[4,5], Prof Dr med; Friedegund Meier[6], Univ Prof Dr med; Frank Friedrich Gellrich[6]; Axel Hauschild[7], Dr med; Lars French[8], Dr med; Justin Gabriel Schlager[8], Dr med; Kamran Ghoreschi[9], Dr med; Tabea Wilhelm[9], Dr med; Heinz Kutzner[10], Dr med; Markus Heppt[11], PD Dr med; Sebastian Haferkamp[12], PD Dr med; Wiebke Sondermann[13], Dr med; Dirk Schadendorf[13], Dr med; Bastian Schilling[14], Dr med; Roman C Maron[1], MSc; Max Schmitt[1], MSc; Tanja Jutzi[1], PhD; Stefan Fröhling[3,15], Prof Dr med; Daniel B Lipka[3,15,16], PD Dr med; Titus Josef Brinker[1], Dr med, MD

[1]Digital Biomarkers for Oncology Group (DBO), National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

[2]Department of Medicine III, RWTH University Hospital Aachen, Aachen, Germany

[3]National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

[4]Department of Dermatology, University Hospital of Mannheim, Mannheim, Germany

[5]Skin Cancer Unit, German Cancer Research Center, Heidelberg, Germany

[6]Skin Cancer Center at the University Cancer Centre and National Center for Tumor Diseases Dresden, Department of Dermatology, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[7]Department of Dermatology, University Hospital of Kiel, Kiel, Germany

[8]Department of Dermatology and Allergology, Ludwig Maximilian University of Munich, Munich, Germany

[9]Department of Dermatology, Venereology and Allergology, Charité – Universitätsmedizin Berlin, Berlin, Germany

[10]Dermatopathology Laboratory, Friedrichshafen, Germany

[11]Department of Dermatology, University Hospital Erlangen, Erlangen, Germany

[12]Department of Dermatology, University Hospital of Regensburg, Regensburg, Germany

[13]Department of Dermatology, University Hospital Essen, Essen, Germany

[14]Department of Dermatology, University Hospital Würzburg, Würzburg, Germany

[15]Translational Cancer Epigenomics, Division of Translational Medical Oncology, German Cancer Research Center, Heidelberg, Germany

[16]Faculty of Medicine, Medical Center, Otto-von-Guericke-University, Magdeburg, Germany

[*]these authors contributed equally

Corresponding Author:
Titus Josef Brinker, Dr med, MD
Digital Biomarkers for Oncology Group (DBO)
National Center for Tumor Diseases (NCT)
German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 460
Heidelberg
Germany
Phone: 49 62213219304
Email: titus.brinker@nct-heidelberg.de

Abstract

Background: Recent years have been witnessing a substantial improvement in the accuracy of skin cancer classification using convolutional neural networks (CNNs). CNNs perform on par with or better than dermatologists with respect to the classification tasks of single images. However, in clinical practice, dermatologists also use other patient data beyond the visual aspects present in a digitized image, further increasing their diagnostic accuracy. Several pilot studies have recently investigated the effects of integrating different subtypes of patient data into CNN-based skin cancer classifiers.

XSL•FO
RenderX

Objective: This systematic review focuses on the current research investigating the impact of merging information from image features and patient data on the performance of CNN-based skin cancer image classification. This study aims to explore the potential in this field of research by evaluating the types of patient data used, the ways in which the nonimage data are encoded and merged with the image features, and the impact of the integration on the classifier performance.

Methods: Google Scholar, PubMed, MEDLINE, and ScienceDirect were screened for peer-reviewed studies published in English that dealt with the integration of patient data within a CNN-based skin cancer classification. The search terms *skin cancer classification*, *convolutional neural network(s)*, *deep learning*, *lesions*, *melanoma*, *metadata*, *clinical information*, and *patient data* were combined.

Results: A total of 11 publications fulfilled the inclusion criteria. All of them reported an overall improvement in different skin lesion classification tasks with patient data integration. The most commonly used patient data were age, sex, and lesion location. The patient data were mostly one-hot encoded. There were differences in the complexity that the encoded patient data were processed with regarding deep learning methods before and after fusing them with the image features for a combined classifier.

Conclusions: This study indicates the potential benefits of integrating patient data into CNN-based diagnostic algorithms. However, how exactly the individual patient data enhance classification performance, especially in the case of multiclass classification problems, is still unclear. Moreover, a substantial fraction of patient data used by dermatologists remains to be analyzed in the context of CNN-based skin cancer classification. Further exploratory analyses in this promising field may optimize patient data integration into CNN-based skin cancer diagnostics for patients' benefits.

(J Med Internet Res 2021;23(7):e20708) doi:10.2196/20708

KEYWORDS

skin cancer classification; convolutional neural networks; patient data

Introduction

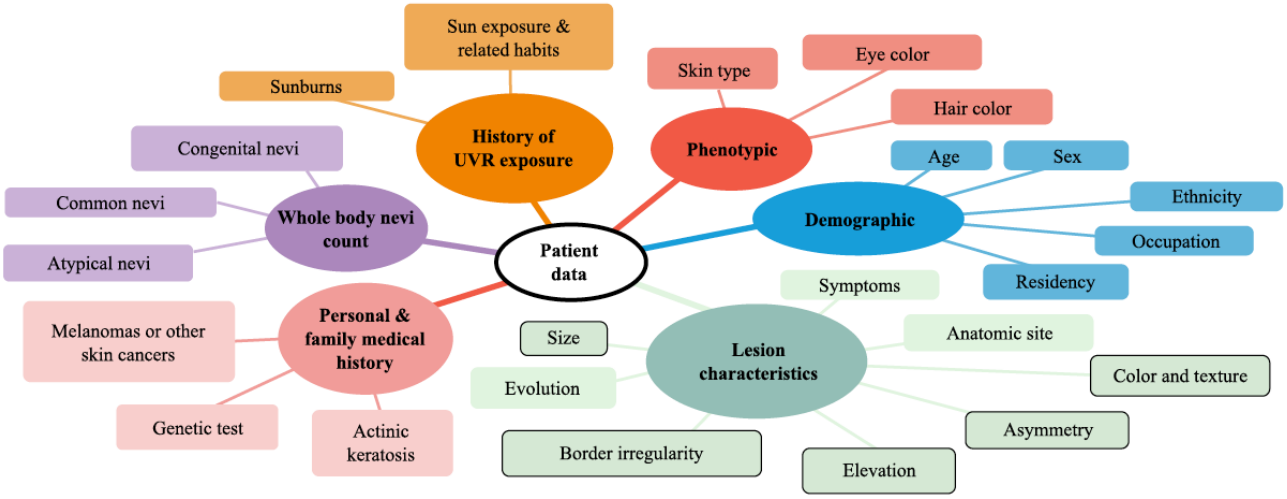
Background

The incidence of skin cancer has been increasing throughout the world, resulting in substantial health and economic burdens [1]. Early detection increases the possibility of curing all types of skin cancers. However, distinguishing benign skin lesions from malignant skin lesions is challenging, even for experienced clinicians [2]. Over the past few years, different digital approaches have been proposed to assist in the detection of skin cancer [3,4]. Convolutional neural networks (CNNs) are the most successful systems for handling image classification problems [5]. Recent publications have reported CNNs that support [6] and outperform [7-9] dermatologists in challenging binary melanoma detection and multiclass skin cancer classification when only taking single images of the skin lesion as input.

However, single-image classification does not reflect the clinical reality. In fact, dermatologists' diagnoses are based on both the visual inspection of a single image and the integration of information from various sources. Figure 1 shows the different types of information that dermatologists may collect from their patients, including information on known risk factors for skin cancer. The corresponding references can be found in Multimedia Appendix 1 [10-32]. The complexity of this figure illustrates the diversity of patient data that can be included in the diagnosis. Roffman et al [33] and Wang et al [34] presented reasonably accurate skin cancer predictions using deep learning methods based exclusively on patient nonimage information. Haenssle et al [35] showed that dermatologists perform somewhat better in a dichotomous skin lesion classification task (benign vs malignant or premalignant) when they were provided with clinical images and textual case information, such as the patient's age, sex, and lesion location, in addition to a dermoscopic image. This raises the question of whether a combination of CNN-based image analysis and patient data might also increase the accuracy of the classifier. A combination of image features and patient data has become a topic of the International Skin Imaging Collaboration challenge in 2019, where a large repository of dermoscopic images including patient data was offered for clinical training and technical research toward automated algorithmic analysis [36].

XSL•FO
RenderX

Figure 1. An overview of patient data considered by dermatologists while diagnosing skin lesions. The framed characteristics in the figure illustrate the fraction of patient data that can potentially be recognized by convolutional neural networks from a single image input. UVR: ultraviolet radiation.



Objective

This review presents the status quo of CNN-based skin lesion classification using image input and patient data. The included studies were analyzed with respect to the amount and type of patient data used for integration, the encoding and fusing techniques, and the reported results. The review also discusses the heterogeneity of the studies that have been conducted so far and points out the potential and challenges of such combined classifiers that should be addressed in the future.

Methods

Search Strategy

Google Scholar, PubMed, MEDLINE, and ScienceDirect were searched for peer-reviewed publications, restricted to human research published in English. The search terms *skin cancer classification*, *convolutional neural network(s)*, *deep learning*, *lesions*, *melanoma*, *metadata*, *clinical information*, and *patient data* were combined.

Study Selection

This review only includes skin lesion classification studies using CNNs that consider both image and patient data. It must be noted that there are a few studies that investigated the incorporation of visual and nonvisual information on skin cancer classification, but did not obtain visual features using deep learning techniques, for example, the studies by Binder et al [37], Alcon et al [38], Cheng et al [39], Liu et al [40], and Rubegni et al [41]. This review exclusively focuses on integrating patient data with the state-of-the-art CNN-based feature extractors. Therefore, the abovementioned studies were not considered in this review.

Study Analysis

The objective of this review is to update practitioners on the status quo approaches toward patient data incorporation into CNN-based skin lesion diagnostics regarding all relevant practical aspects.

Type and Amount of Patient Data

The goal is to achieve better performance of the CNN-based classifier by integrating new information that cannot be extracted from a digitized image. Various types of patient data have been shown to assist dermatologists. *Key question:* Which and how many different types of patient data have been tested for CNN-based classification?

Encoding and Fusing Techniques

A CNN-based classifier extracts various visual features from a digitized image as the basis for its diagnosis. Patient data are nonimage data and are mostly provided as numbers or strings in tables. The patient data can be classified in a dichotomous fashion (presence of the feature: yes or no), fall into several discrete categories (eg, Fitzpatrick skin type), or be continuous (eg, patient age). This may require different, carefully chosen encoding and fusing techniques. Moreover, the weight attributed to patient data in comparison with image features can strongly influence how the different features contribute toward the decision making of the system. *Key questions:* What are the encoding and fusing techniques applied in the studies? Do the studies focus on the quantitative relationship between image and nonimage features?

Reported Study Results

This review aims to summarize the recent findings regarding the impact of patient data on the performance of CNN-based classifiers. *Key questions:* What is the classification task? Is it a binary or multiclass problem? Which skin lesions should be distinguished? How is the influence of individual and/or combined patient data documented? In the case of multiclass classification, is the impact also shown for each single class of skin lesion individually?

Applied Performance Metrics

The included publications reported different statistical metrics as the study end points. If the classes in the test set are approximately equally distributed, then accuracy is a frequently used performance metric, where the total number of correctly predicted samples is divided by the total number of samples in

XSL•FO
RenderX

the test set. In binary classification problems with a positive and a negative class, sensitivity and specificity are further common study end points, especially if there is an imbalance between the samples of both classes. Sensitivity was determined only on the basis of the actual positive samples in the test set. It is calculated by counting the correctly classified positive samples by the total number of positive samples. In contrast, specificity was determined based on actual negative samples in the test set. Here, the correctly classified negative samples were divided by the total number of negative samples. While using a CNN, the sensitivity and specificity depend on the selected cutoff value. If the output of the neural network is greater than the cutoff value, the input is assigned to the positive class, and if it is below that value, then the input is assigned to the negative class. Thus, this value represents a central parameter for the trade-off between sensitivity and specificity. A decrease in the threshold value leads to an increase in the sensitivity with a simultaneous decrease in specificity and vice versa. The dependence of the cutoff value of the specificity and sensitivity of the two metrics is shown in the receiver operating characteristic curve. Here, the sensitivity is plotted against the false-positive rate (1−specificity) in a diagram for each possible cutoff value. The area under the receiver operating characteristic curve was used as an integral performance measure for the algorithms.

Results

Classification Tasks

A total of 11 publications fulfilling the inclusion criteria are summarized in Table 1. The studies were very heterogeneous with respect to CNN architecture, classification task, including image and patient data, data augmentation (if reported), and fusion techniques, rendering a meaningful direct comparison very difficult. A total of 5 studies dealt with binary classifications with the end point of either dichotomous melanoma or basal cell carcinoma (BCC) classification or with the end point to distinguish malignant from benign lesions in general. The remaining 6 studies were classified between 5 and 8 different skin diseases or lesions. As malignant lesions, melanoma (6 studies), BCC (6 studies), and squamous cell carcinoma (SCC; 3 studies) were included. In addition, these studies differentiated among the following benign lesions: melanocytic nevus (NV; 6 studies), benign keratosis-like lesions (BKLs; 4 studies), dermatofibroma (3 studies), vascular lesions (VASCs; 3 studies), actinic keratosis (AK; 2 studies), and seborrheic keratosis (SK; 2 studies). Moreover, merged groups comprising AK and intraepithelial carcinoma or Bowen disease (2 studies) and dermatofibroma, lentigo, melanosis, miscellaneous, and VASCs (1 study) were used in some of the studies.

XSL•FO
RenderX

Table 1. Summary table.

Study	Patient data types	Result (without/with)	Classification task	CNN[a] architecture	Data set	Samples, n
Bonechi et al [42]	4 types: age, sex, location, and presence of melanocytic cells	Accuracy: 0.8344/0.8834	Binary: benign or malignant (MEL[b], BCC[c], SCC[d])	ResNet50	ISIC[e]	5405
Chin et al [43]	5 types: age; sex; size; how long it existed; changes in size, color, or shape including bleeding and itching	Accuracy: 0.84/0.92	Binary: low risk or high risk for MEL	DenseNet121	Own	5289
Gonzalez-Diaz [44]	2 types: age and sex	Accuracy: 0.848/0.859	Binary: MEL yes or no	ResNet50	2017 ISBI[f] challenge+interactive atlas of dermoscopy [45]+ISIC	6302
Gessert et al [46]	3 types: age, sex, and location	Sensitivity: 0.725/0.742; specificity: data not available	8 classes: MEL, NV[g], BCC, AK[h], BKL[i], DF[j], VASC[k], SCC	EfficientNets	ISIC (HAM10000 [47], BCN_2000 [48], MSK [49])+7-point data set [50]	27,665
Kawahara et al [50]	3 types: sex, location, and elevation	Sensitivity: 0.527/0.604; specificity: 0.902/0.910	5 classes: MEL, BCC, NV, MISC[l], SK[m]	Inception V3	7-point data set	808
Kharazmi et al [51]	5 types: age, sex, location, size, and elevation	Accuracy: 0.847/0.911	Binary: BCC yes or no	Convolutional filters of learned kernel weights from a sparse autoencoder	Own	1199
Li et al [52]	3 types: age, sex, and location	Sensitivity: 0.8544/0.8764; specificity: data not available	7 classes: NV, MEL, BKL, BCC, AKIEC[n], VASC, DF	SENet154	ISIC 2018 data set	10,015
Pacheco and Krohling [53]	8 types: age, location, lesion itches, bleeds or has bled, pain, recently increased, changed its pattern, and elevation	Accuracy: 0.671/0.788	6 Classes: BCC, SCC, AK, SK, MEL, NV	ResNet50	Own	1612
Ruiz-Castilla et al [54]	3 types: age, sex, and size	Accuracy: 0.61/0.85	Binary: MEL yes or no	Shallow network with 2 convolutional layers	ISIC	300
Sriwong et al [55]	3 types: age, sex, and location	Accuracy: 0.7929/0.8039	7 classes: AKIEC, BCC, BKL, DF, MEL, NV, VASC	AlexNet	HAM10000	16,720
Yap et al [56]	3 types: age, sex, and location	Mean average precision: 0.726/0.729; Accuracy: 0.721/0.720	5 classes: BCC, SCC, MEL, BKL, NV	ResNet50	ILSVRC[o] 2015 [57]+own	2917 (only testing)

[a]CNN: convolutional neural network (most of the studies had the goal of investigating the usefulness of the presented fusion technique independently of the convolutional neural network architecture and, therefore, often showed the performance of the fusion with multiple architectures; we included only the best-performing architecture).

[b]MEL: melanoma.

[c]BCC: basal cell carcinoma.

[d]SCC: squamous cell carcinoma.

[e]ISIC: International Skin Imaging Collaboration.

[f]ISBI: International Symposium on Biomedical Imaging [49].

[g]NV: melanocytic nevus.

[h]AK: actinic keratosis.

[i]BKL: benign keratosis-like lesion.

[j]DF: dermatofibroma.

[k]VASC: vascular lesion.

XSL•FO
RenderX

[l]MISC: summary of dermatofibroma, lentigo, melanosis, miscellaneous, and vascular lesion.

[m]SK: seborrheic keratosis.

[n]AKIEC: actinic keratosis and intraepithelial carcinoma.

[o]ILSVRC: ImageNet Large Scale Visual Recognition Challenge.

Types and Amount of Patient Data

Most of the studies included three types of patient data (7/11, 64%). Compared with the diversity of potentially useful patient data illustrated in Figure 1, only a few types of patient data were considered. The most commonly included types of data were patient's age and sex (studies: 10/11, 91%). Only Kawahara et al [50] and Pacheco and Krohling [53] did not consider age and sex, respectively. The third most commonly considered feature was lesion location (studies: 8/11, 73%). Elevation and lesion size were considered in 27% (3/11) of studies. Chin et al [43] and Pacheco and Krohling [53] included statements about symptoms such as itching, bleeding or pain. In addition, they tracked the lesion's evolution by documenting whether the lesion increased in size or changed its shape. Furthermore, Bonechi et al [42] considered the presence of melanocytic cells as an additional potentially relevant feature.

Encoding

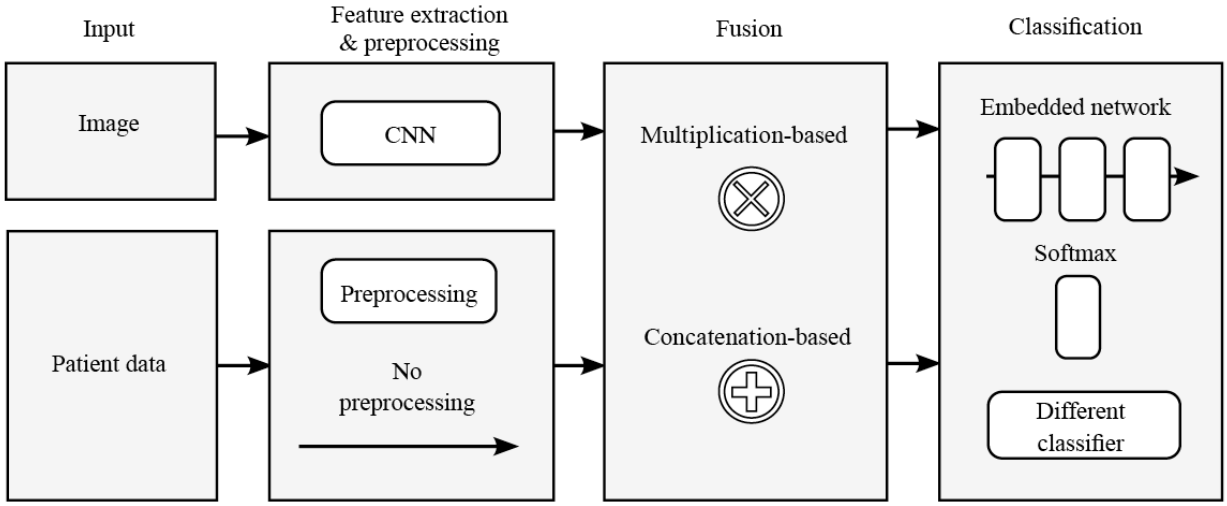
The means of choice to encode the patient data was one-hot encoding in most cases. One-hot encoding is one way to encode several discrete classes with a string of bits, where exactly one value in the string of bits encoding one class is assigned 1 and all others are assigned 0 (eg, *melanoma*=010; *BCC*=100; *NV*=001). Different techniques were used for continuous parameters such as the patient's age. One-hot encoding is only possible after discretizing the continuous range, which was performed by Bonechi et al [42], who divided the age ranging from 0 to 95 in the sections of 5 years. Gessert et al [46] tested numerical against one-hot encoding and found the former to be superior. Li et al [52] normalized the age in the range between 0 and 1 and represented its information using only one value.

As patient data are rarely documented in a standardized way, dealing with missing values is an essential skill that requires the algorithm to be proficient. However, only 18% (2/11) of publications went into detail on how they dealt with missing values. Gessert et al [46] suggested a negative fixed value for missing data, whereas Li et al [52] used the more common approach to fill in missing values with average values for continuous data and the most frequent values for discrete patient data.

Fusing Technique

Figure 2 illustrates the main function blocks in which the studies vary with respect to the fusing techniques.

Figure 2. Overview of the different fusing techniques in the main function blocks of the combined classifier. CNN: convolutional neural network.



The fusing techniques differ in the way they actively weigh the image and patient data. In 82% (9/11) of studies, a concatenation-based fusion was applied, that is, the feature vector extracted from the images was enlarged by attaching the encoded patient data. In this case, weighting is achieved by defining the ratio between the number of features originating from the image and the patient data input. Common CNN architectures extract 1024, 2048, or even more features from the image input. In most studies, the authors decided to reduce the image features before concatenating them with patient data. In only 27% (3/11) of studies, the authors provided sufficient information on this point and revealed a considerable variance in the ratio of image features to patient data: 112 to 28 [53], 128 to 80 [42], and 2048 or 2×2048 to 11 [56]. However, reducing the image features should be done with care, as it is accompanied by a loss of information. Only Pacheco and Krohling [53] reported the effect of changing the ratio and proved its strong influence on the classification performance. A totally different weighting approach was introduced by Li et al [52]. This approach used a multiplication-based fusion. Inspired by the squeeze-and-excitation operation of a SENet network [58], the authors used patient data to control the importance of each image feature channel at the last convolutional layer. Thus, the network was able to focus on

XSL•FO
RenderX

specific parts of the image feature based on patient data. The authors determined the multiplication-based fusion to be superior to the concatenation-based approach in multiple network architectures.

In addition, the studies vary in the extent to which deep learning methods were applied to the patient data before fusing or on the combined feature vector after fusing it with the image data. Sriwong et al [55] applied no further deep learning methods but used a separate support vector machine for classification, which received the image features extracted by the CNN and the encoded patient data as input. Gonzalez-Diaz [44] used a separate support vector machine for the patient data to generate a probabilistic output, which was factorized with the output of the CNN-based classifier to provide the final diagnosis of the system. Because of the end-to-end training of a neural network, a direct fusion within the CNN architecture was used in most studies. Kharazmi et al [51] simply added patient data before the last classifying softmax layer. More complex deep learning methods were applied by Gessert et al [46] before concatenation or using an embedded network comprising multiple fully connected layers after concatenation as presented by Pacheco and Krohling [53] and Yap et al [56].

Reported Study Results

As summarized in Table 1, all but one study reported a considerable improvement in the classification performance when patient data were used in addition to image analysis. However, the authors consistently emphasized that patient data are only a support source and the image features clearly provide the main evidence [53,56]. Yap et al [56], who considered the features of age, sex, and lesion location for a multiclass problem (BCC, SCC, melanoma, BKL, NV), concluded in the discussion that their incorporation of patient data showed only a slight but not significant improvement in accuracy and recommended testing different features, such as nevus count, proportion of atypical nevi, and history of melanoma.

Although 5 studies reported results for binary classification tasks, 55% (6/11) of studies dealt with a multiclass classification problem, distinguishing between up to 8 different skin diseases, and revealed insights on how the use of patient data influences the classification performance for an individual type of skin lesion. Table 2 shows which individual classifications benefitted from the integration of patient data and whether this was achieved at the expense of others. Among the 6 studies, Gessert et al [46], Sriwong et al [55], and Li et al [52] dealt with comparable classification tasks and used the same patient data (age, sex, and location). All 3 studies identified improvements in the classification of BKL and dermatofibroma. Li et al [52] even listed an absolute increase in the sensitivity for dermatofibroma of approximately 20% (from 63.56% to 84.55%). It must be stated critically that the authors failed to mention the corresponding specificity, which makes it difficult to draw reliable conclusions. Furthermore, Table 2 shows that the improvements may go along with the degradation of classification performance for other lesion types [52,55] or that the improvement of sensitivity for one class may be paralleled by a decrease in specificity, as shown clearly by the results of Gessert et al [46]. In contrast, Kawahara et al [50] and Pacheco and Krohling [53] used different patient data (eg, elevation of the lesion) and reported an increase in sensitivity and specificity in almost all classes. Unfortunately, a deeper insight into the study results of Yap et al [56] was not possible because the confusion matrix, including the classification performance of the single-lesion types, was not legible.

XSL•FO
RenderX

Table 2. Influence of included patient data on the classification performance of the single skin diseases or lesions[a].

Study, patient data, and metric	Skin disease										
	MEL[b]	NV[c]	BCC[d]	SCC[e]	AK[f]	AKIEC[g]	BKL[h]	DF[i]	VASC[j]	MISC[k]	SK[l]
Gessert et al [46]: age, sex, location											
AUC[m]	+[n]	(+/−)[o]	–[p]	−	+	X[q]	+	+	−	X	X
Sensitivity	−	−	−	−	−	X	−	−	−	X	X
Specificity	+	+	+	+	+	X	+	+	+	X	X
Sriwong et al [55]: age, sex, location											
Sensitivity	+	−	+	X	X	−	+	+	−	X	X
Specificity	−	+	−	X	X	+	+	+/−	+	X	X
Li et al [52]: age, sex, location											
Sensitivity	−	−	+	X	X	−	+	+	+	X	X
Kawahara et al [50]: sex, location, elevation											
Sensitivity	+	+	+	X	X	X	X	X	X	+	+
Specificity	+	+	+/−	X	X	X	X	X	X	+	+
Pacheco and Krohling [53]: age, location, itches, bleeds, pain, increased, changed, elevation											
Sensitivity	+	+	+	+	+	X	X	X	X	X	+
Specificity	+	+	+	−	+	X	X	X	X	X	+

[a]The study of Yap et al [56] is excluded because the confusion matrix was not legible. It must be noticed that there are some combinations where the outcome deteriorates by including patient data.

[b]MEL: melanoma.

[c]NV: Melanocytic nevus.

[d]BCC: basal cell carcinoma.

[e]SCC: squamous cell carcinoma.

[f]AK: Actinic keratosis.

[g]AKIEC: actinic keratosis and intraepithelial carcinoma.

[h]BKL: benign keratosis-like lesion.

[i]DF: dermatofibroma.

[j]VASC: vascular lesion.

[k]MISC: miscellaneous and vascular lesion.

[l]SK: seborrheic keratosis.

[m]AUC: area under the curve.

[n]Indicates improvement compared with classification performance without patient data.

[o]Indicates no change compared with classification performance without patient data.

[p]Indicates degradation compared with classification performance without patient data.

[q]This implies that the lesion type was not considered in the classification task of the study.

In total, 36% (4/11) of studies analyzed the influence of the used patient data on the classification performance in a more differentiated way. They showed the impact of either individual patient data or special combinations of patient data on classification performance, thereby providing a more detailed insight into the contribution of individual patient data.

As the only ones, Pacheco and Krohling [53] performed an exploratory analysis of the patient data within the used data set before observing the classification of the CNN. The authors considered eight types of patient data (age, location, lesion itches, lesion bleeds, lesion hurts [“pain”], recent increase in size, changed shape or pattern, and elevation) and six different types of skin lesions (BCC, SCC, AK, SK, melanoma, and NV). The exploratory analysis suggested that the patient data parameters such as “bleeding” and “pain” were suitable to differentiate between pigmented (NV, melanoma, and SK) and nonpigmented lesions (AK, BCC, and SCC), whereas the patient data parameters such as “changed its pattern” and “elevation” helped to identify melanomas. As “pain” was always denied in the case of AK, this feature seemed to be a promising discriminator. The analyzed patient data for SCC and BCC were very similar; therefore, no improvement was expected for these 2 skin diseases because of the integration of patient data. The classification results confirmed the exploratory analysis because the classifications of AK, melanoma, NV, and SK improved

XSL•FO
RenderX

when patient data were incorporated into the classifiers, whereas the performance for BCC and SCC remained almost the same.

Li et al [52] considered three types of patient data (age, sex, and lesion location) and 7 skin lesion types (NV, melanoma, BKL, BCC, AK and intraepithelial carcinoma, VASC, and dermatofibroma). The study showed the overall classification performance for all possible combinations of patient data. The integration of parameter "location" resulted in the best classification performance, individually. The combination of patient data parameters of "age" and "location" provided the best result overall, whereas the parameter "sex" decreased performance upon integration. The authors concluded that the rare diseases of VASC and dermatofibroma are more location specific, whereas none of the skin diseases in question occur preferentially in men or women. Therefore, the authors recommended the use of "location" and the avoidance of "sex" in the combined classifier.

Sriwong et al [55] addressed the same problem as that of Li et al [52]. However, their study only analyzed some combinations of patient data (age, age+sex, and age+sex+location). The best overall result was achieved by incorporating the combination of "age," "sex," and "location." Contrary to Li et al [52], the study yielded the largest improvement for the feature "age." Although adding "sex" did not show a considerable improvement, additionally adding "location" increased the performance slightly. The authors stated that the information of "sex" and "location" is more powerful when used in combination, thereby confirming statements in related studies [40,59].

Bonechi et al [42] considered four types of patient data individually (age, sex, location, and presence of melanocytic cells) for a binary classification (malignant yes or no). Unfortunately, the analysis results have not been reported in detail, but the authors reported the parameter "presence of melanocytic cells" to be the most informative.

Discussion

Principal Findings

Although the main evidence for a good diagnosis is still provided by the image input, all 11 publications indicate a possible benefit of integrating patient data in CNN classifiers, as illustrated in Table 1. This corresponds with the results of other approaches that combine visual and nonvisual features for skin lesion classification [37-41], thereby suggesting it as a promising avenue of research. However, publication bias favoring studies with positive results cannot be excluded.

One focus of further research into combined CNN-based classifiers should be to render its classification process transparent, easy to understand, and applicable in a clinical setting. The 11 studies published so far have dealt with these aspects only marginally. Therefore, these issues need to be addressed in future studies to reliably reveal the potential of integrating patient data.

Reproducibility, Comparability, and Generalization

No objective benchmarks exist in the field of integrating patient data into CNN-based classifiers. The heterogeneity of the studies conducted so far is substantial. This applies to the number and types of skin diseases or lesions to be classified, databases and data augmentation, CNN architectures, patient data, and fusion techniques. These aspects have a great influence on the way that the algorithm learns to diagnose the lesions in question and render it very difficult to reproduce and compare the approaches and results externally and independently. A way to solve this would be the more extensive use of external and publicly available data sets to objectively optimize the classification accuracy in an experimental setting. This needs to be done systematically in preparation for clinical trials that will be required to prove the algorithm's generalizability and applicability in the clinic. In addition, the best way to handle missing data needs to be addressed.

Transparency and Explainability

All presented studies lack an investigation of the impact of patient data individually and in combination on single-lesion classes. Both the fusion method and weight attributed to the patient data in addition to the biological significance itself may substantially influence the classification results. Further research should be dedicated to explaining the mechanisms by which the incorporation of these factors contributes to the decision making of the CNN-based combined classifier to render the results more transparent.

Call for Extensive Exploration Analysis

As shown in Figure 1, a diversity of patient data has been shown to be useful in a clinical setting and could be considered for diagnosis by CNN-based classifiers as well. So far, researchers have mostly used patient data that are readily available and/or routinely recorded, such as age, sex, and lesion location. However, readily available factors may not be the best choice. For instance, sex was included in 91% (10/11) of studies, but was stated to be of minimal benefit for the classification task if investigated in detail. Regarding the results of studies considering patient data besides these three factors, the results indicate that the integration of other patient data may be more promising [39,42,53]. Studies analyzing the risk factors for skin cancer so far demonstrated that patient data can be helpful in distinguishing skin lesions in binary classification tasks. Corresponding studies are available for differentiating between melanoma and nonmelanoma skin cancer [10,60] and for distinguishing between BCC and SCC [61]. Patient data, such as the skin type (I, II, III vs IV), the count of atypical nevi (>4 vs none) and common nevi (>100 vs 0-4) are well-established criteria for melanoma risk [11,12]. To our knowledge, no extensive study has analyzed significant correlations among individual or combinations of these types of patient data and the improvement of multiclass problems as considered in this review. An extensive exploration analysis in this field would help to choose patient data suitable for the considered classification task. Following the study of Haenssle et al [35], it would further be of interest to note which type of patient data influences the clinician's decision the most. Studies comparing the benefit of specific patient data integration in the artificial

XSL•FO
RenderX

intelligence system versus the clinician's decision and, therefore, pointing out the opportunities of human-algorithm integration systems should be the subject of future research.

Conclusions

All 11 studies published so far indicate that the integration of patient data into CNN-based skin lesion classifiers may improve classification accuracy. The studies mainly used patient data that were routinely recorded (age, sex, and lesion location). Regarding the technical details, the main differences in the presented approaches occur in the fusing techniques. Further research should be dedicated to systematically evaluating the impact of incorporation of individual and combined patient data into CNN-based classifiers to show its benefit reproducibly and transparently and to pave the way for the translation of these combined classifiers into the clinic.

Acknowledgments

JH, A Heckler, EKH, and TJB contributed to the concept and design of the study. JH identified and analyzed studies with contributions from A Heckler and EKH. TJB oversaw the study, critically reviewed and edited the manuscript, and gave final approval. JNK, JSU, FM, FFG, A Hauschild, LF, JGS, KG, TW, HK, MH, SH, WS, DS, BS, RCM, MS, TJ, SF, and DBL substantially contributed to the conception and design, provided critical review and commentary on the draft manuscript, and approved the final version. All of the authors guaranteed the integrity and accuracy of this study. This research is funded by the Federal Ministry of Health in Germany (Skin Classification Project; grant holder: TJB).

Conflicts of Interest

A Hauschild reports clinical trial support, speaker's honoraria, and consultancy fees from the following companies: Amgen, Bristol Myers Squibb (BMS), Merck Serono, Merck Sharp & Dohme (MSD), Novartis, Oncosec, Philogen, Pierre Fabre, Provectus, Regeneron, Roche, OncoSec, Sanofi Genzyme, and Sun Pharma (outside the submitted work). BS reports advisory roles for or has received honoraria from Pierre Fabre Pharmaceuticals, Incyte, Novartis, Roche, BMS, and MSD; research funding from BMS, Pierre Fabre Pharmaceuticals and MSD; and travel support from Novartis, Roche, BMS, Pierre Fabre Pharmaceuticals, and Amgen outside the submitted work. FM has received travel support, speaker's fees, and/or advisor's honoraria from Novartis, Roche, BMS, MSD, and Pierre Fabre and research funding from Novartis and Roche outside the submitted work. JSU is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GSK, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, Roche, outside the submitted work. SH reports advisory roles for or has received honoraria from Pierre Fabre Pharmaceuticals, Novartis, Roche, BMS, Amgen, and MSD outside the submitted work. TJB reports owning a company that develops mobile apps (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg). WS received travel expenses for attending meetings and/or (speaker) honoraria from Abbvie, Almirall, Bristol Myers Squibb, Celgene, Janssen, LEO Pharma, Lilly, MSD, Novartis, Pfizer, Roche, Sanofi Genzyme, and UCB outside the submitted work.

Multimedia Appendix 1
Relevant references for the overview of the patient data illustrated.
[DOCX File , 19 KB - jmir_v23i7e20708_app1.docx]

References

1. Schadendorf D, van Akkooi AC, Berking C, Griewank KG, Gutzmer R, Hauschild A, et al. Melanoma. Lancet 2018 Sep;392(10151):971-984. [doi: 10.1016/s0140-6736(18)31559-9]
2. Brinker TJ, Hekler A, Hauschild A, Berking C, Schilling B, Enk AH, et al. Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. Eur J Cancer 2019 Apr;111:30-37 [FREE Full text] [doi: 10.1016/j.ejca.2018.12.016] [Medline: 30802784]
3. Dick V, Sinz C, Mittlböck M, Kittler H, Tschandl P. Accuracy of computer-aided diagnosis of melanoma: a meta-analysis. JAMA Dermatol 2019 Nov 01;155(11):1291-1299 [FREE Full text] [doi: 10.1001/jamadermatol.2019.1375] [Medline: 31215969]
4. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin cancer classification using convolutional neural networks: systematic review. J Med Internet Res 2018 Oct 17;20(10):e11936 [FREE Full text] [doi: 10.2196/11936] [Medline: 30333097]
5. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput 2017 Sep;29(9):2352-2449. [doi: 10.1162/neco_a_00990]
6. Han SS, Park I, Chang SE, Lim W, Kim MS, Park GH, et al. Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. J Invest Dermatol 2020 Sep;140(9):1753-1761 [FREE Full text] [doi: 10.1016/j.jid.2020.01.019] [Medline: 32243882]
7. Maron R, Weichenthal M, Utikal J, Hekler A, Berking C, Hauschild A, Collabrators. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer 2019 Sep;119:57-65 [FREE Full text] [doi: 10.1016/j.ejca.2019.06.013] [Medline: 31419752]

XSL•FO
RenderX

8. Brinker T, Hekler A, Enk A, Klode J, Hauschild A, Berking C, Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer 2019 May;113:47-54 [FREE Full text] [doi: 10.1016/j.ejca.2019.04.001] [Medline: 30981091]
9. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol 2019 Jul;20(7):938-947. [doi: 10.1016/s1470-2045(19)30333-x]
10. Krensel M, Petersen J, Mohr P, Weishaupt C, Augustin J, Schäfer I. Estimating prevalence and incidence of skin cancer in Germany. J Dtsch Dermatol Ges 2019 Dec 30;17(12):1239-1249. [doi: 10.1111/ddg.14002] [Medline: 31885171]
11. S3 Leitlinie Prävention Hautkrebs, AWMF Registernummer: 032/052OL. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF). URL: http://leitlinienprogrammonkologie.de/Leitlinien.7.0.html [accessed 2021-06-07]
12. Bataille V, Bishop J, Sasieni P, Swerdlow A, Pinney E, Griffiths K, et al. Risk of cutaneous melanoma in relation to the numbers, types and sites of naevi: a case-control study. Br J Cancer 1996 Jun;73(12):1605-1611 [FREE Full text] [doi: 10.1038/bjc.1996.302] [Medline: 8664138]
13. Consensus Development Panel. National institutes of health summary of the consensus development conference on sunlight, ultraviolet radiation, and the skin. J Am Acad Dermatol 1991 Apr;24(4):608-612. [doi: 10.1016/s0190-9622(08)80159-4]
14. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Solar and ultraviolet radiation. In: International Agency for Research on Cancer. Geneva: World Health Organization; 1992.
15. Malignes Melanom der Haut : ICD-10 C43. Krebsdaten. URL: https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Melanom/melanom_node.html [accessed 2020-04-05]
16. Gandini S, Sera F, Cattaruzza MS, Pasquini P, Abeni D, Boyle P, et al. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. Eur J Cancer 2005 Jan;41(1):28-44 [FREE Full text] [doi: 10.1016/j.ejca.2004.10.015] [Medline: 15617989]
17. Marcil I, Stern RS. Risk of developing a subsequent nonmelanoma skin cancer in patients with a history of nonmelanoma skin cancer: a critical review of the literature and meta-analysis. Arch Dermatol 2000 Dec 01;136(12):1524-1530. [doi: 10.1001/archderm.136.12.1524] [Medline: 11115165]
18. Salasche SJ. Epidemiology of actinic keratoses and squamous cell carcinoma. J Am Acad Dermatol 2000 Jan;42(1 Pt 2):4-7 [FREE Full text] [doi: 10.1067/mjd.2000.103342] [Medline: 10607349]
19. Evans T, Boonchai W, Shanley S, Smyth I, Gillies S, Georgas K, et al. The spectrum of patched mutations in a collection of Australian basal cell carcinomas. Hum Mutat 2000 Jul;16(1):43-48. [doi: 10.1002/1098-1004(200007)16:1<43::aid-humu8>3.0.co;2-7]
20. Lock-Andersen J, Drzewiecki KT, Wulf HC. Eye and hair colour, skin type and constitutive skin pigmentation as risk factors for basal cell carcinoma and cutaneous malignant melanoma. A Danish case-control study. Acta Derm Venereol 1999 Jan 18;79(1):74-80 [FREE Full text] [doi: 10.1080/000155599750011778] [Medline: 10086866]
21. Titus-Ernstoff L, Perry AE, Spencer SK, Gibson JJ, Cole BF, Ernstoff MS. Pigmentary characteristics and moles in relation to melanoma risk. Int J Cancer 2005 Aug 10;116(1):144-149 [FREE Full text] [doi: 10.1002/ijc.21001] [Medline: 15761869]
22. Veierød MB, Weiderpass E, Thörn M, Hansson J, Lund E, Armstrong B, et al. A prospective study of pigmentation, sun exposure, and risk of cutaneous malignant melanoma in women. J Natl Cancer Inst 2003 Oct 15;95(20):1530-1538. [doi: 10.1093/jnci/djg075] [Medline: 14559875]
23. Lasithiotakis K, Krüger-Krasagakis S, Ioannidou D, Pediaditis I, Tosca A. Epidemiological differences for cutaneous melanoma in a relatively dark-skinned Caucasian population with chronic sun exposure. Eur J Cancer 2004 Nov;40(16):2502-2507 [FREE Full text] [doi: 10.1016/j.ejca.2004.06.032] [Medline: 15519526]
24. Fargnoli MC, Piccolo D, Altobelli E, Formicone F, Chimenti S, Peris K. Constitutional and environmental risk factors for cutaneous melanoma in an Italian population. A case-control study. Melanoma Res 2004 Apr;14(2):151-157 [FREE Full text] [doi: 10.1097/00008390-200404000-00013] [Medline: 15057047]
25. Naldi L, Imberti GL, Parazzini F, Gallus S, La Vecchia C. Pigmentary traits, modalities of sun reaction, history of sunburns, and melanocytic nevi as risk factors for cutaneous malignant melanoma in the Italian population. Cancer 2000 Jun 15;88(12):2703-2710. [doi: 10.1002/1097-0142(20000615)88:12<2703::aid-cncr8>3.0.co;2-q]
26. Ródenas JM, Delgado-Rodríguez M, Herranz MT, Tercedor J, Serrano S. Sun exposure, pigmentary traits, and risk of cutaneous malignant melanoma: a case-control study in a Mediterranean population. Cancer Causes Control 1996 Mar;7(2):275-283 [FREE Full text] [doi: 10.1007/bf00051303]
27. Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, et al. The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. J Am Acad Dermatol 1994 Apr;30(4):551-559. [doi: 10.1016/s0190-9622(94)70061-3]
28. Marzuka A, Book S. Basal cell carcinoma: pathogenesis, epidemiology, clinical features, diagnosis, histopathology, and management. Yale J Biol Med 2015 Jun;88(2):167-179 [FREE Full text] [Medline: 26029015]
29. Gordon R. Skin cancer: an overview of epidemiology and risk factors. Semin Oncol Nurs 2013 Aug;29(3):160-169. [doi: 10.1016/j.soncn.2013.06.002] [Medline: 23958214]

XSL•FO
RenderX

30. Scrivener Y, Grosshans E, Cribier B. Variations of basal cell carcinomas according to gender, age, location and histopathological subtype. Br J Dermatol 2002 Jul;147(1):41-47. [doi: 10.1046/j.1365-2133.2002.04804.x] [Medline: 12100183]
31. Augustin J, Kis A, Sorbe C, Schäfer I, Augustin M. Epidemiology of skin cancer in the German population: impact of socioeconomic and geographic factors. J Eur Acad Dermatol Venereol 2018 Nov 17;32(11):1906-1913. [doi: 10.1111/jdv.14990] [Medline: 29633375]
32. Pion I, Rigel D, Garfinkel L, Silverman M, Kopf A. Occupation and the risk of malignant melanoma. Cancer 1995 Jan 15;75(2 Suppl):637-644. [doi: 10.1002/1097-0142(19950115)75:2+<637::aid-cncr2820751404>3.0.co;2-#] [Medline: 7804988]
33. Roffman D, Hart G, Girardi M, Ko CJ, Deng J. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. Sci Rep 2018 Jan 26;8(1):1701 [FREE Full text] [doi: 10.1038/s41598-018-19907-9] [Medline: 29374196]
34. Wang H, Wang Y, Liang C, Li Y. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. JAMA Dermatol 2019 Sep 04:2335. [doi: 10.1001/jamadermatol.2019.2335] [Medline: 31483437]
35. Haenssle H, Fink C, Toberer F, Winkler J, Stolz W, Deinlein T, Reader Study Level I and Level II Groups. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. Ann Oncol 2020 Jan;31(1):137-143 [FREE Full text] [doi: 10.1016/j.annonc.2019.10.013] [Medline: 31912788]
36. ISIC Challenge 2019. URL: https://challenge2019.isic-archive.com/ [accessed 2020-05-11]
37. Binder M, Kittler H, Dreiseitl S, Ganster H, Wolff K, Pehamberger H. Computer-aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process. Melanoma Res 2000 Dec;10(6):556-561. [doi: 10.1097/00008390-200012000-00007] [Medline: 11198477]
38. Alcon JF, Ciuhu C, ten Kate W, Heinrich A, Uzunbajakava N, Krekels G, et al. Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. IEEE J Sel Top Signal Process 2009 Feb;3(1):14-25. [doi: 10.1109/jstsp.2008.2011156]
39. Cheng B, Stanley RJ, Stoecker WV, Stricklin SM, Hinton KA, Nguyen TK, et al. Analysis of clinical and dermoscopic features for basal cell carcinoma neural network classification. Skin Res Technol 2013 Feb 22;19(1):217-222 [FREE Full text] [doi: 10.1111/j.1600-0846.2012.00630.x] [Medline: 22724561]
40. Liu Z, Sun J, Smith M, Smith L, Warr R. Incorporating clinical metadata with digital image features for automated identification of cutaneous melanoma. Br J Dermatol 2013 Nov 31;169(5):1034-1040. [doi: 10.1111/bjd.12550] [Medline: 23902335]
41. Rubegni P, Feci L, Nami N, Burroni M, Taddeucci P, Miracco C, et al. Computer-assisted melanoma diagnosis: a new integrated system. Melanoma Res 2015 Dec;25(6):537-542. [doi: 10.1097/CMR.0000000000000209] [Medline: 26426763]
42. Bonechi S, Bianchini M, Bongini P, Ciano G, Giacomini G, Rosai R, et al. Fusion of visual and anamnestic data for the classification of skin lesions with deep learning. In: Cristani M, Prati A, Lanz O, Messelodi S, Sebe N, editors. New Trends in Image Analysis and Processing - ICIAP 2019. Switzerland: Springer; 2019:211-219.
43. Chin Y, Hou Z, Lee M, Chu H, Wang H, Lin Y, et al. A patient-oriented, general-practitioner-level, deep-learning-based cutaneous pigmented lesion risk classifier on a smartphone. Br J Dermatol 2020 Jun 06;182(6):1498-1500. [doi: 10.1111/bjd.18859] [Medline: 31907926]
44. Gonzalez-Diaz I. DermaKNet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. IEEE J Biomed Health Inform 2019 Mar;23(2):547-559. [doi: 10.1109/jbhi.2018.2806962]
45. Argenziano G, Soyer HP, De Giorgio V, Piccolo D, Carli P, Delfinoo M, et al. Interactive Atlas of Dermoscopy. Milan, Italy: Edra Medical Publishing & New Media; 2000:1-208.
46. Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefer A. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. MethodsX 2020;7:100864 [FREE Full text] [doi: 10.1016/j.mex.2020.100864] [Medline: 32292713]
47. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018 Aug 14;5:180161 [FREE Full text] [doi: 10.1038/sdata.2018.161] [Medline: 30106392]
48. Combalia M, Codella NC, Rotemberg V, Helba B, Vilaplana V, Reiter O, et al. BCN20000: Dermoscopic lesions in the wild. arXiv:1908.02288. 2019. URL: https://arxiv.org/abs/1908.02288 [accessed 2021-06-14]
49. Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) 2018:168-172. [doi: 10.1109/ISBI.2018.8363547]
50. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE J Biomed Health Inform 2019 Mar;23(2):538-546 [FREE Full text] [doi: 10.1109/jbhi.2018.2824327]

XSL•FO
RenderX

51. Kharazmi P, Kalia S, Lui H, Wang ZJ, Lee TK. A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile. Skin Res Technol 2018 May 22;24(2):256-264. [doi: 10.1111/srt.12422] [Medline: 29057507]
52. Li W, Zhuang J, Wang R, Zhang J, Zheng W. Fusing metadata and dermoscopy images for skin disease diagnosis. In: Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020 Presented at: IEEE 17th International Symposium on Biomedical Imaging (ISBI); April 3-7, 2020; Iowa City, IA, USA. [doi: 10.1109/isbi45749.2020.9098645]
53. Pacheco AG, Krohling RA. The impact of patient clinical information on automated skin cancer detection. Comput Biol Med 2020 Jan;116:103545. [doi: 10.1016/j.compbiomed.2019.103545] [Medline: 31760271]
54. Ruiz-Castilla J, Rangel-Cortes J, García-Lamont F, Trueba-Espinosa A. CNN and metadata for classification of benign and malignant melanomas. In: Intelligent Computing Theories and Application, ICIC 2019, Lecture Notes in Computer Science, Vol 11644. Switzerland: Springer; 2019:569-579.
55. Sriwong K, Bunrit S, Kerdprasop K, Kerdprasop N. Dermatological classification using deep learning of skin image and patient background knowledge. Int J Mach Learn Comput 2019 Dec;9(6):862-867 [FREE Full text] [doi: 10.18178/ijmlc.2019.9.6.884]
56. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. Exp Dermatol 2018 Nov 27;27(11):1261-1267. [doi: 10.1111/exd.13777] [Medline: 30187575]
57. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. IJCV, 2015. URL: https://image-net.org/challenges/LSVRC/2015/index.php [accessed 2021-06-14]
58. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 2020 Aug;42(8):2011-2023 [FREE Full text] [doi: 10.1109/TPAMI.2019.2913372] [Medline: 31034408]
59. Visconti A, Ribero S, Sanna M, Spector TD, Bataille V, Falchi M. Body site-specific genetic effects influence naevus count distribution in women. Pigment Cell Melanoma Res 2020 Mar 25;33(2):326-333. [doi: 10.1111/pcmr.12820] [Medline: 31403758]
60. Jerant A, Johnson J, Sheridan C, Caffrey T. Early detection and treatment of skin cancer. Am Fam Physician 2000 Jul 15;62(2):357-68, 75 [FREE Full text] [Medline: 10929700]
61. Firnhaber J. Diagnosis and treatment of basal cell and squamous cell carcinoma. Am Fam Physician 2012 Jul 15;86(2):161-168 [FREE Full text] [Medline: 22962928]

Abbreviations

AK: actinic keratosis
BCC: basal cell carcinoma
BKL: benign keratosis-like lesion
BMS: Bristol Myers Squibb
CNN: convolutional neural network
MSD: Merck Sharp & Dohme
NV: melanocytic nevus
SCC: squamous cell carcinoma
SK: seborrheic keratosis
VASC: vascular lesion

Edited by G Eysenbach; submitted 27.05.20; peer-reviewed by R Kaczmarczyk, L Li; comments to author 20.10.20; revised version received 29.10.20; accepted 13.04.21; published 02.07.21.

Please cite as:
Höhn J, Hekler A, Krieghoff-Henning E, Kather JN, Utikal JS, Meier F, Gellrich FF, Hauschild A, French L, Schlager JG, Ghoreschi K, Wilhelm T, Kutzner H, Heppt M, Haferkamp S, Sondermann W, Schadendorf D, Schilling B, Maron RC, Schmitt M, Jutzi T, Fröhling S, Lipka DB, Brinker TJ
Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review
J Med Internet Res 2021;23(7):e20708
URL: https://www.jmir.org/2021/7/e20708
doi:10.2196/20708
PMID:34255646

©Julia Höhn, Achim Hekler, Eva Krieghoff-Henning, Jakob Nikolas Kather, Jochen Sven Utikal, Friedegund Meier, Frank Friedrich Gellrich, Axel Hauschild, Lars French, Justin Gabriel Schlager, Kamran Ghoreschi, Tabea Wilhelm, Heinz Kutzner,

XSL•FO
RenderX

Markus Heppt, Sebastian Haferkamp, Wiebke Sondermann, Dirk Schadendorf, Bastian Schilling, Roman C Maron, Max Schmitt, Tanja Jutzi, Stefan Fröhling, Daniel B Lipka, Titus Josef Brinker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis

Ardalan Mirzaei[1*], BPharm, MPhil, GCertEdStud (Higher Ed); Parisa Aslani[1*], BPharm (Hons), MSc, PhD, GCertEdStud (Higher Ed); Edward Joseph Luca[2], BA, MBA; Carl Richard Schneider[1*], BN, BPharm (Hons), PhD, PGCert (Higher Ed)

[1]The University of Sydney School of Pharmacy, Faculty of Medicine and Health, The University of Sydney, Australia

[2]The University of Sydney Library, The University of Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Ardalan Mirzaei, BPharm, MPhil, GCertEdStud (Higher Ed)
The University of Sydney School of Pharmacy
Faculty of Medicine and Health
Pharmacy and Bank Building A15
The University of Sydney, 2006
Australia
Phone: 61 286275225
Email: ardalan.mirzaei@sydney.edu.au

Abstract

Background: People engage in health information–seeking behavior to support health outcomes, and being able to predict such behavior can inform the development of interventions to guide effective health information seeking. Obtaining a comprehensive list of the predictors of health information–seeking behavior through a systematic search of the literature and exploring the interrelationship of these predictors are critical first steps in this process.

Objective: This study aims to identify significant predictors of health information–seeking behavior in the primary literature, develop a common taxonomy for these predictors, and identify the evolution of the concerned research field.

Methods: A systematic search of PsycINFO, Scopus, and PubMed was conducted for all years up to and including December 10, 2019. Quantitative studies identifying significant predictors of health information–seeking behavior were included. Information seeking was broadly defined and not restricted to any source of health information. Data extraction of significant predictors was performed by 2 authors, and network analysis was conducted to observe the relationships between predictors with time.

Results: A total of 9549 articles were retrieved, and after the screening, 344 studies were retained for analysis. A total of 1595 significant predictors were identified. These predictors were categorized into 67 predictor categories, with the most central predictors being age, education, gender, health condition, and financial income. With time, the interrelationship of predictors in the network became denser, with the growth of new predictor grouping reaching saturation (1 new predictor identified) in the past 7 years, despite increasing publication rates.

Conclusions: A common taxonomy was developed to classify 67 significant predictors of health information–seeking behavior. A time-aggregated network method was developed to track the evolution of the research field, showing the maturation of new predictor terms and an increase in primary studies reporting multiple significant predictors of health information–seeking behavior. The literature has evolved with a decreased characterization of novel predictors of health information–seeking behavior. In contrast, we identified a parallel increase in the complexity of predicting health information–seeking behavior, with an increase in the literature describing multiple significant predictors.

(J Med Internet Res 2021;23(7):e21680) doi:10.2196/21680

KEYWORDS

information seeking; network analysis; health; review; temporal analysis; mobile phone

XSL•FO
RenderX

Introduction

Background

Health information seeking has been defined as "the ways in which individuals obtain information, including information about their health, health promotion activities, risks to one's health, and illness" [1]. A consumer's health information–seeking behavior has the potential to influence the process and outcomes related to coping or adjusting to an illness or condition [1].

The conceptualization of health information–seeking behavior has evolved since Lenz first defined it in 1984, which identified 2 dimensions of health information–seeking behavior: extent and method [2]. Lambert and Loiselle's comprehensive review [1] of the concept of health information–seeking behavior found definitions such as actions or behaviors used to obtain knowledge [3], clarify or confirm knowledge [4], satisfy a query [5], identify information sources [6,7], or demonstrate a coping strategy [8,9]. Other known models from the information science perspective include the Comprehensive Model of Information Seeking, which looks at information carrier characteristics, antecedents, and information-seeking actions [10] and the book by Case in 2002 about the research on information-seeking needs and behaviors [11]. A recent paper by Zimmerman and Shaw describes health information–seeking behavior as an *umbrella term* for many forms of information seeking, such as "direct seeking, information monitoring or browsing, and the passive receipt of information" [12]. Alternatively, people's passive receipt of health information has been defined separately as *information scanning* [13], where people may not be active in their search but are still receptive enough to receive information. Thus, although the concept of health information–seeking behavior has existed for more than 30 years, there remains a lack of consensus on its definition and model theories.

Despite these inconsistencies, previous models and theories commonly describe health information–seeking behavior as involving the action of seeking out information, regardless of where it comes from, how it is sought, or why it is sought [1,14]. Therefore, predictors of health information–seeking behavior can be described as the variables affecting the actions of seeking out information. These predictors can be contextual, such as the environment of an individual or their social networks, and personal, such as sociodemographic characteristics, health status, or internal beliefs. There may also be predictors of persistence in health information–seeking behavior, such as satisficing. Satisficing relates to decision making "through which an individual decides when an alternative approach or solution is sufficient to meet the individuals' desired goals rather than pursue the perfect approach" [15]. In the context of information seeking, it is choosing whether it is worth the cost or effort of continuing to search or whether already acquired information suffices.

Factors that influence why people engage in information seeking include the content of the information, which information sources or channels are frequently used, their credibility, and the barriers they may pose to seeking information [1,16]. A meta-analytic review by Chang and Huang quantified 7 predictors of health information–seeking behavior; however, not all of the predictors, such as behaviors (adherence) and beliefs, were included in their review [17]. Predicting peoples' behavior for health information seeking requires understanding the predictors and their significance and magnitude on information-seeking behavior.

We define *significant predictors* as those shown through empirical research to have a direct effect rather than an association (correlation) to health information seeking. An example of this direct effect is age. Nelissen et al [18] showed that an increase in age led to increased cancer information seeking. An example of an association is the relationship between patient-physician interaction and information-seeking behavior. In this case, it is unclear whether information seeking leads to better patient-physician interaction (an outcome of health information–seeking behavior) or whether better physician-patient interaction leads to increased health information–seeking behavior (a predictor of health information–seeking behavior). Although associations between predictors and health information–seeking behavior may have statistical significance in some empirical studies, knowing the direction of the effect from predictor to variable is more informative. Furthermore, although qualitative research can provide a foundation for identifying predictors of health information–seeking behavior, the ability to quantify the effect size allows for comparison of individual predictors' relative importance.

A comprehensive list of predictors of health information–seeking behavior provides researchers with a focus on identifying new significant predictors or examining the relationship and effect of new interventions. Predictors of health information–seeking behavior can support researchers, clinicians, private institutions, and public health initiatives in optimizing information-related interactions between themselves and consumers, leading to a more positive health care management experience [1,16].

Consistent classification of terms is the first step in formulating a comprehensive list of health information–seeking behavior predictors. For instance, there is concern that the terms *race* and *ethnicity* have been used synonymously in health research despite being separate constructs [19]. Thus, how predictors of health information–seeking behavior are defined in one study may not necessarily be consistent with another.

Anker et al [16] compiled a comprehensive list of predictors for health information–seeking behavior a decade ago through a systematic search of the literature. They extracted and reported on the methods and measures used in health information–seeking behavior research. However, there are 2 critical shortcomings of this review. First, Anker et al [20] restricted their definition of health information–seeking behavior to an active process, in accordance with Niederdeppe's definition of health information–seeking behavior. Consequently, predictors for the nonactive acquisition of health information were not identified [20]. Second, their search strategy was restricted to a single database (PsycINFO), with the justification that health information–seeking behavior is a social psychological construct instead of a medical construct. PsycINFO has focused subject

XSL•FO
RenderX

areas, and it is possible that other health information–seeking behavior researchers may have published in journals not indexed in PsycINFO.

Importantly, there is a need for an updated review to account for the evolution of information seeking as a result of the rapid emergence and dominance of mobile digital information technology. The use of the internet has been increasing in the past three decades [21]. Advances in technologies such as smartphones have led to increased availability and access to the internet. Since 2011, smartphone ownership by the American population has increased from 35% to 81% in 2019, with 96% of the population owning a cellular device [22]. Similarly, the use of smartphones has led to greater access to the internet, with the exclusive use of smartphones for internet access doubling from a reported 17.5% in 2013 to 37% of the American population in 2019 [21,23]. A smartphone user is estimated to spend a daily average of 2.6 hours on their device [24]. Although the internet has become a common source of health information [25], how the influence of the internet has modified predictors of health information–seeking behavior throughout time is yet to be well characterized. Previous studies have compared sources of information used by people as part of health information–seeking behavior; however, most studies have only compared the findings from the early 2000s with those from the early 2000s [26-28]. Such comparison studies report that internet use was not a predictor of information-seeking behavior, yet Huerta et al [26] reported an increase in internet use, especially in older age groups. In contrast, Li et al [28] performed a hierarchical regression analysis comparing 2002 and 2012 cohorts from the Pew Database to examine changes in health information–seeking behavior on the web. They identified internet access as a predictor for health information decline with time. The authors hypothesized that this could be partly because of the increase in misinformation and rise in smartphone use, resulting in increased accessibility of the internet as a source of health information in the United States. The extent to which these changes have affected predictors of health information–seeking behavior as a whole and across information sources and settings has yet to be reviewed.

Objectives

This review aims to identify predictors of health information–seeking behavior, as reported in the primary literature, and explore the relationships between predictors with time. The specific objectives are as follows:

- identify significant predictors of health information–seeking behavior in the primary literature;
- develop a common taxonomy for predictors of health information–seeking behavior;
- identify the evolution of the health information–seeking behavior research field using quantitative studies.

Methods

Selection Criteria

The following section outlines the inclusion and exclusion criteria for this review.

Types of Participants

The papers included defined participants as health consumers or caregivers. The intent of the search for information was important: a health consumer searches for information for their own self or treatment, as opposed to a health professional who may search for information to provide therapy. Caregivers were also included, as they were in a nontherapeutic relationship with the health consumer.

Papers in which the participants were health students (university or college) or simulation studies in which the participants sought information prospectively in hypothetical scenarios were excluded. Students studying a health-related degree were removed if in their health-related disciplines, they were searching for information for their future role as health professionals.

Types of Studies

Quantitative studies were eligible for inclusion in this study. Relevant study designs included experimental and epidemiological studies, including randomized controlled trials, nonrandomized controlled trials, quasi-experimental studies, before-and-after studies, prospective and retrospective cohort studies, case-control studies, and analytical cross-sectional studies.

An article was included if it reported the significance level of a predictor. That is, the study showed that a certain *predictor* was significant in causing *information seeking* rather than associations. Significance was determined by either *P* values <.05 or in the case of odds ratio if the CI did not cross 1. Articles were also included if further work demonstrated a causation or effect size, which was often shown through logistic or linear regression, confirmatory factor analysis, or structural equation modeling.

Studies were excluded if they did not have a quantitative focus. These included qualitative studies such as focus groups, semistructured interviews, mixed methods studies (with no quantitative component), and content analysis of websites or interviews.

Search Procedures

Search Limits

Papers published in English up to December 10, 2019, since database inception were considered for inclusion. No data range was applied. Participants' information seeking was not restricted to any source, and all sources (eg, web, health care practitioner) were included.

Databases

The databases searched were PsycINFO, Scopus, and PubMed. Scopus is considered the largest abstract and citation database of peer-reviewed literature and incorporates the results from Embase and MEDLINE [29].

Search Terms or Phrases

The keywords used were the following:

XSL•FO
RenderX

Health OR *Drug* OR *Medicine*; AND *Information Seeking* OR *Information Behavior* OR *Information Search* or *Satisficing* (Tables S1-S3 in Multimedia Appendix 1 for full syntax).

Screening

Two authors (AM and EJL) independently screened a random split of articles for inclusion by title and abstract using the selection criteria. Pilot tests were conducted to calibrate the screening process before the records were split.

Data Extraction

Each included paper was counted as a data source for the extraction; 2 authors (AM and EJL) independently extracted the following variables: year of publication, country of the study, participant recruitment, disease states, theories used, and significant predictors. Significant predictors were those variables for which direct effects were reported (not correlations) on health information–seeking behavior and provided significance with either *P* values <.05 or CIs (in the case of odds ratio) that did not cross 1. Significant relative predictors of health information–seeking behavior were also extracted from studies reporting comparative health information–seeking behavior between 2 or more groups. Any uncertainty associated with extraction was mediated by a third author (CRS), who performed an audit of approximately 10% (34/344) of the screened articles and extracted variables.

Analysis

Content Analysis

The first author (AM) analyzed the variables and categorized the significant predictors into emerging categories. Individual predictors identified in the individual papers were extracted, and an iterative process of clustering was undertaken by 2 authors (AM and CRS); the 2 authors reached a consensus for terms and categories. The categories consisted of identifying similarity between predictor terms: where predictors were the same, they were categorized together; where predictors were similar with a common definition, they were categorized together; and finally, where there was no common definition, but predictors were described similarly in text, they were categorized together [30].

Predictor Frequency

As part of the content analysis, a *word frequency* analysis was performed. In this case, the *words* chosen were identified-predictor terms. Examining the predictor term frequency assists in analyzing the strength and importance of a predictor with regard to other terms [31-34]. Each predictor extracted into a category was counted as 1 for that article. Multiple predictors, if categorized, were categorized as 1. For example, if the article reported “Age 25-30” and “Age 65-70” as significant factors, then they would be categorized as *age*; however, they would contribute only 1 to the *age* category for that article instead of 2. The total predictors were then reported, and the predictor frequency was used to develop the network structure for network analysis.

Network Analysis

Network analysis has been used in previous systematic reviews to identify relationships among authors of the included papers [35] or in a health context to compare drug treatments [36]. Traditionally, quantitative data from a systematic review are pooled via meta-analysis, which requires homogeneous data. Network analysis allows the examination of relationships among heterogeneous entities [37]. A network analysis was conducted to observe the relationships between the predictor terms. This method can help identify nodes (or predictor terms) connected to other nodes and show the relationships between terms in the literature [38,39].

The weights of the nodes were based on the frequency of the predictor, whereas their size was based on the number of articles mentioning the predictor term. An analysis of changes throughout time was undertaken to compare the networks of articles before and after 2008 (articles dated up to December 31, 2008)—while Pew Research started reporting smartphone ownership in 2011 [22], the iPhone was the beginning of a new phone era [40]. The year 2008 was chosen to distinguish between the availability of smartphones following the introduction of the iPhone in 2007, allowing for the uptake of the device to have begun [41]. Accordingly, in 2008, global mobile broadband subscriptions overtook fixed broadband subscriptions [42]. Time-based comparisons of temporal and atemporal network features were observed using time-varying networks. Such an approach has been used in ecology, transport, and social media [43-45].

The co-occurrence of individual predictors within an article was calculated based on the predictor frequency. Each individual predictor term was connected bidirectionally to another predictor from the same article. Each connection adds a weight of 1 to the edge. Edges were formed where a pair of predictor terms was mentioned together in an article. The visualization was created using R software (R Foundation for Statistical Computing), with the code available on GitHub [46]. The co-occurrence of predictors and visualization of the network was created using the igraph package, a software package used for network visualizations between different objects on a network map [47].

The number of nodes and edges, along with modularity, were captured to compare the different networks. Modularity measures the clustered communities of nodes, which is how the nodes cluster together, forming a community group of nodes distant to another community group of nodes. The full setup and parameters are available in GitHub [46]. See Multimedia Appendix 1 for further methods [38,39,48,49]

Results

Search Results

The literature search process is illustrated in Figure 1. From the 2 databases, a total of 9549 articles were retrieved, of which 2866 were duplicates. Following deduplication, title and abstract screening was performed, followed by full-text screening. A total of 344 papers were included in the final analysis. The results of the categorized predictors are reported in Table S4 in

XSL•FO
RenderX

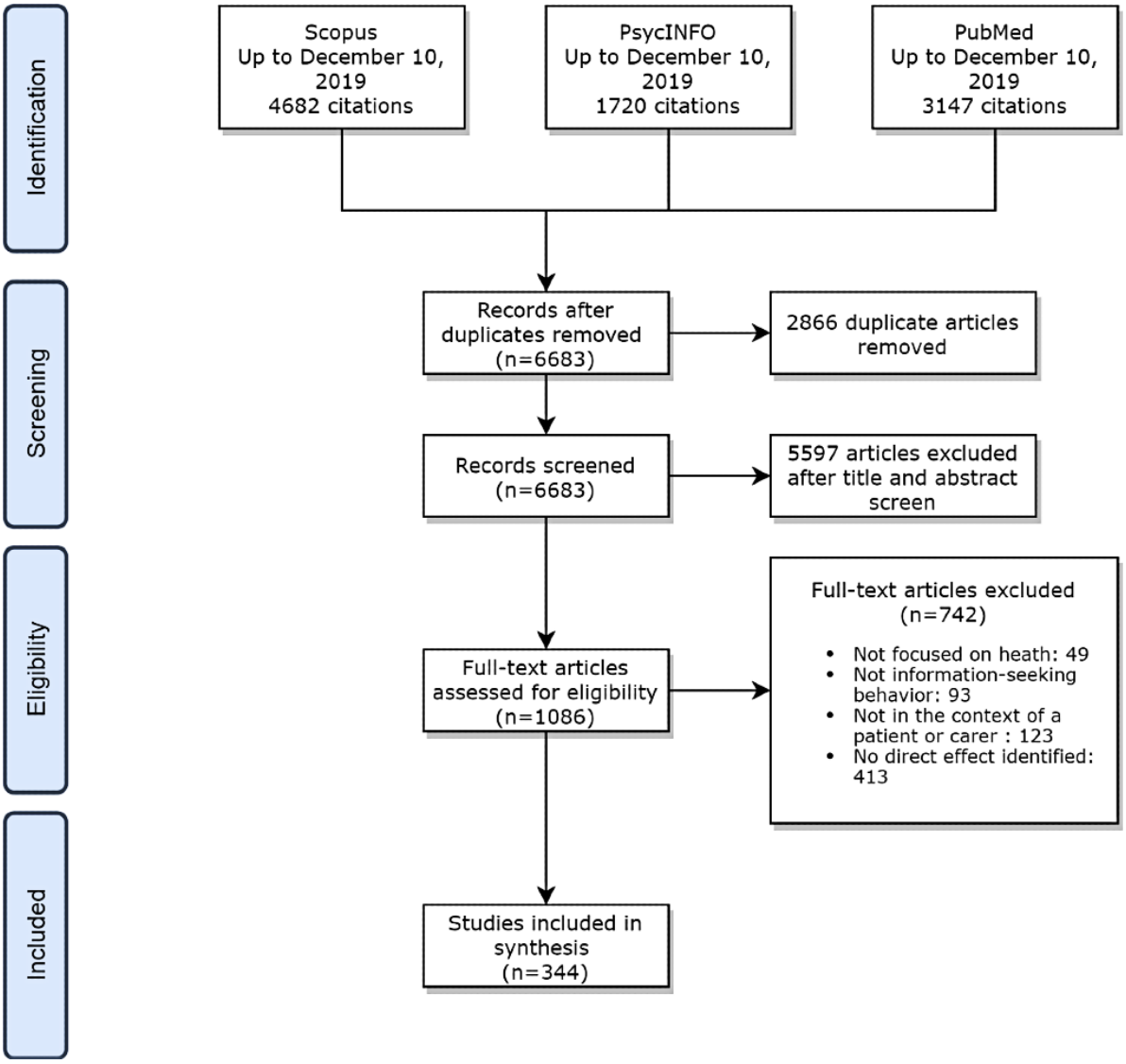
Multimedia Appendix 1. The included articles contained papers published between 1993 and 2019, with a peak publication year in 2019 (Figure 2).

Most of the studies were conducted in the United States (n=202); 26 articles reported studies from China, 12 from Australia and South Korea, and 9 from the United Kingdom and Germany.

In 203 articles, participants were recruited specifically for the study. However, another source of participants was using existing databases of respondents such as the Health Information Trends Survey (HINTS; n=65 papers), The Pew Research Center's Internet & American Life Project (n=9), and the Pennsylvania Cancer Registry (n=8).

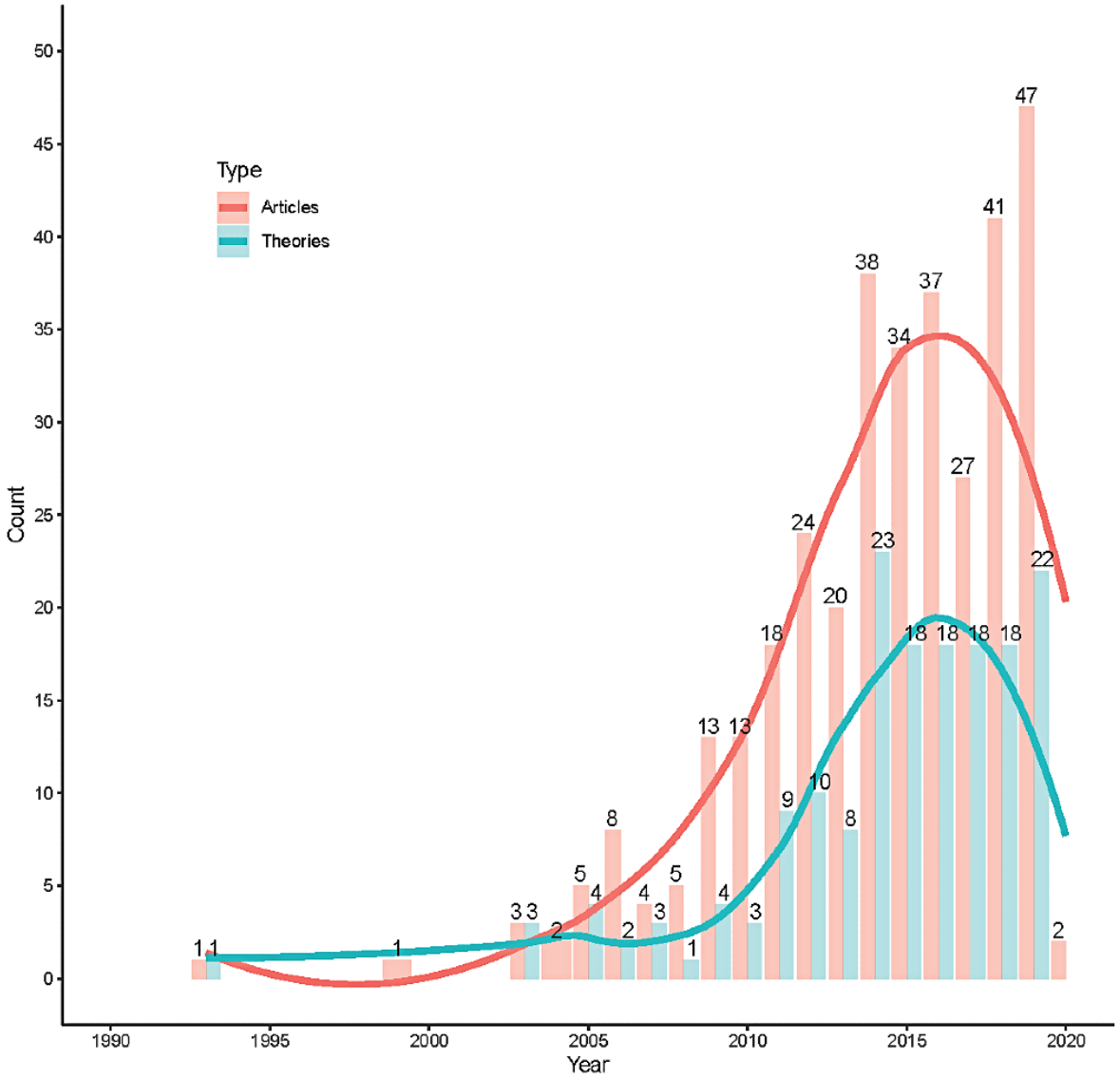
Participants were predominantly seeking information for chronic diseases, with cancer being the most studied condition (n=76).

Figure 1. Flowchart for article inclusion. Reasons for exclusion do not sum to the number of excluded articles because some articles overlapped in their reasons for exclusion.



XSL•FO
RenderX

Figure 2. Total articles published each year and the number of articles that used a theoretically grounded approach.



Content Analysis

Fewer than half of the papers (n=167) were underpinned by a theory or model. As shown in Figure 2, there was an increase in the number of papers with a theoretical underpinning until 2014, at which point a plateau developed from 2015 to 2018. In 2019, there was a second peak of articles with theoretical underpinnings. However, less than half of the papers published in 2019 were supported by a theory or model (Figure 2).

A total of 1595 nonunique, significant predictors were identified. Table S4 in Multimedia Appendix 1 lists the predictor categories. Predictors were classified into 68 categories (1 category was labeled *unclassifiable/other*; unclassifiable predictors were not carried forward for network analysis.). The categories were further grouped into sociodemographic, health, information source, information content, and affective predictor groups. Sociodemographic variables of education (n=160), age (n=156), and gender (n=120) were the most commonly reported significant predictor categories, followed by the health-related predictor categories of health condition (n=87). A noted increase in the number of predictors reported in the literature began in 2005 and peaked in 2019.

Network Analysis

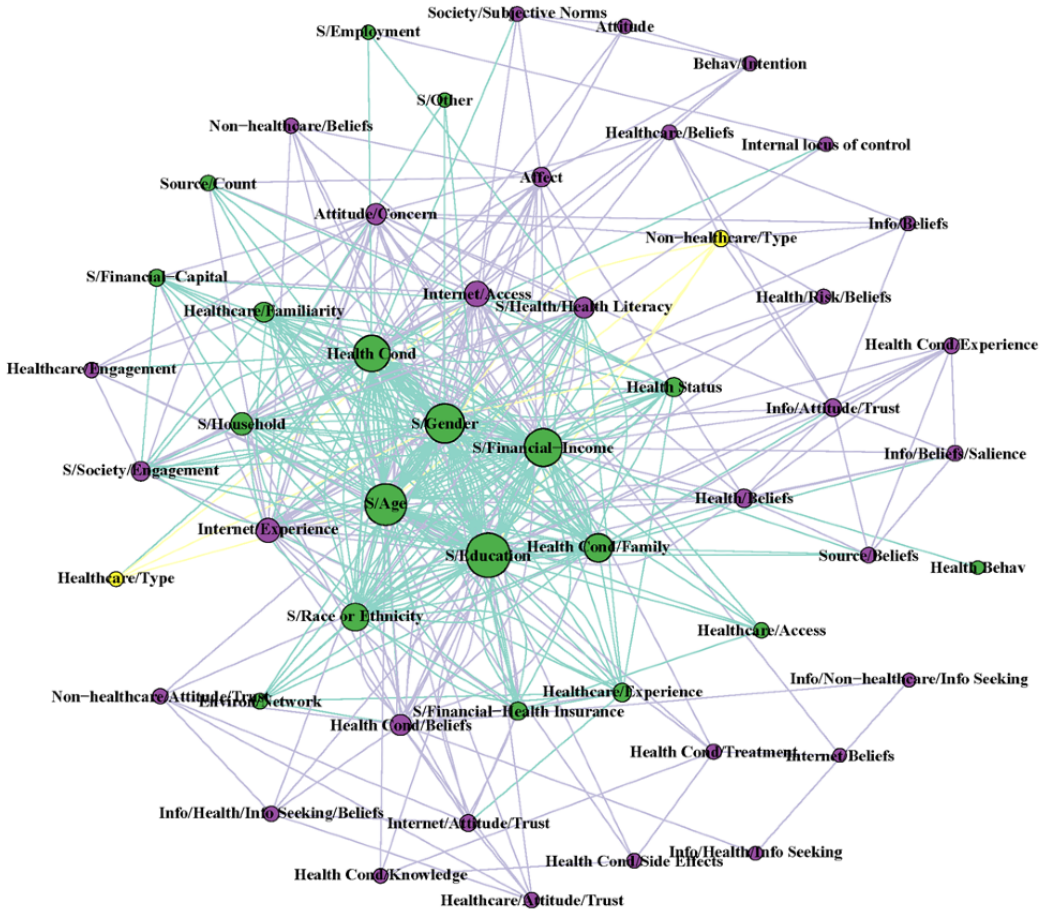
The complete network with all terms and years combined resulted in 67 nodes (*other* node not included) and 4128 edge connections. The modularity of the groups revealed 3 clusters. The largest group of variables was predominately composed of psychosocial predictors (n=41). The second-largest group was sociodemographic predictors (n=22), followed by a third group that did not have any strong focus on any particular grouping (n=4). However, most sociodemographic variables (age, education, gender) had the greatest eigenvector centrality before other variables (health condition and financial income), thus appearing in the center of the network (Table S5 in Multimedia Appendix 1).

The network statistics reported in Table S5 of Multimedia Appendix 1 as well as in Figures 3 and 4 show a difference in structural characteristics before and after 2008. After 2008, only 15 new nodes were identified, with no new nodes identified after 2014 (Table S4 in Multimedia Appendix 1). There was a 7.8 times greater number of edges in the post-2008 network than in the pre-2008 network. The combination of an increased number of edges and a limited increase in nodes resulted in a

XSL•FO
RenderX

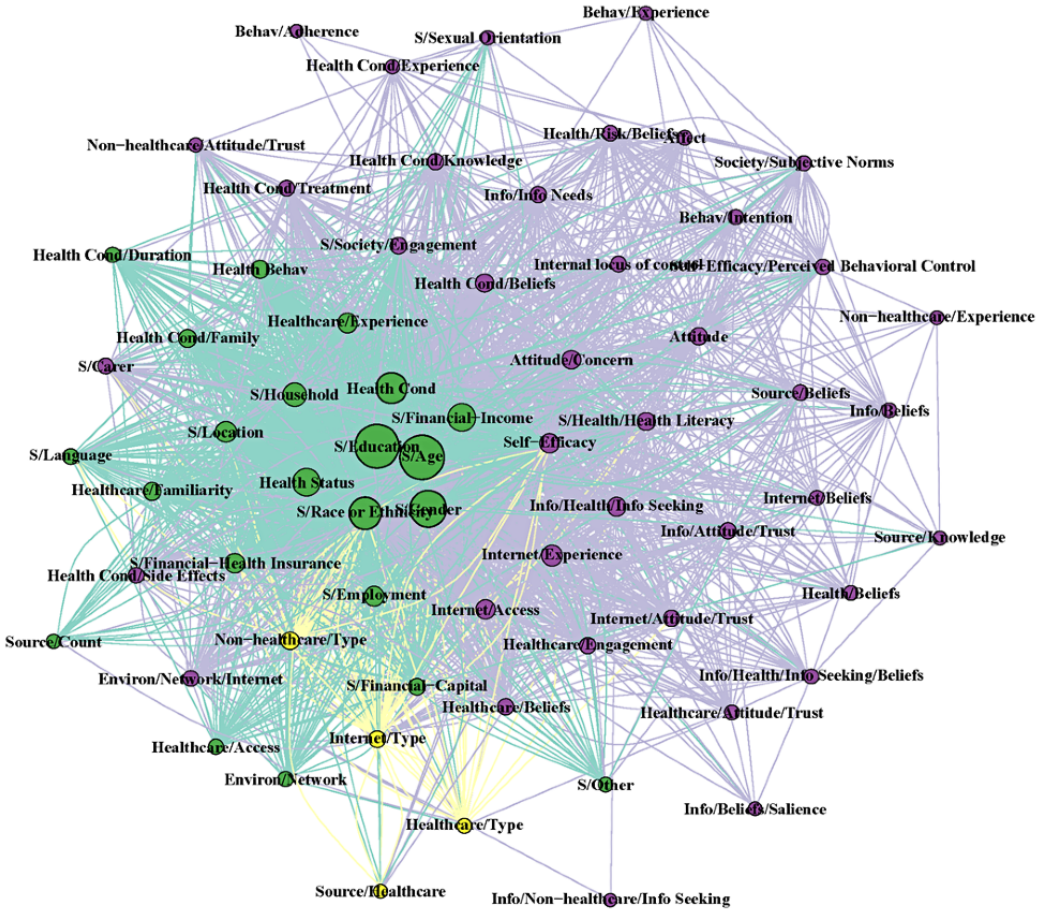
more connected network after 2008, with the average number of adjacent edges to each node (mean degree of the nodes) increased by 6.1 times compared with before 2008. Age, education, gender, and health condition were the nodes with the greatest degree of centrality, indicating the greatest influence on adjacent nodes (Figure 5). Modularity was greatest before 2008; it decreased to 2 in the post-2008 period, indicating tighter communities of nodes clustering together, with all years compared being 3. This tighter clustering is because of the greater co-occurrence of the predictor terms being researched. That is, individual articles reported more significant predictors than previous articles. A sensitivity analysis was conducted to compare networks before and after 2014 (the most recent new node), which confirmed the dynamic shifts in network statistics after 2008 (Table S5 in Multimedia Appendix 1).

Figure 3. Fruchterman-Reingold layout algorithm of the network analysis comparing the pre- and including 2008 network models with color coding according to the modularity group membership of the complete model. The repository is available on GitHub [46].



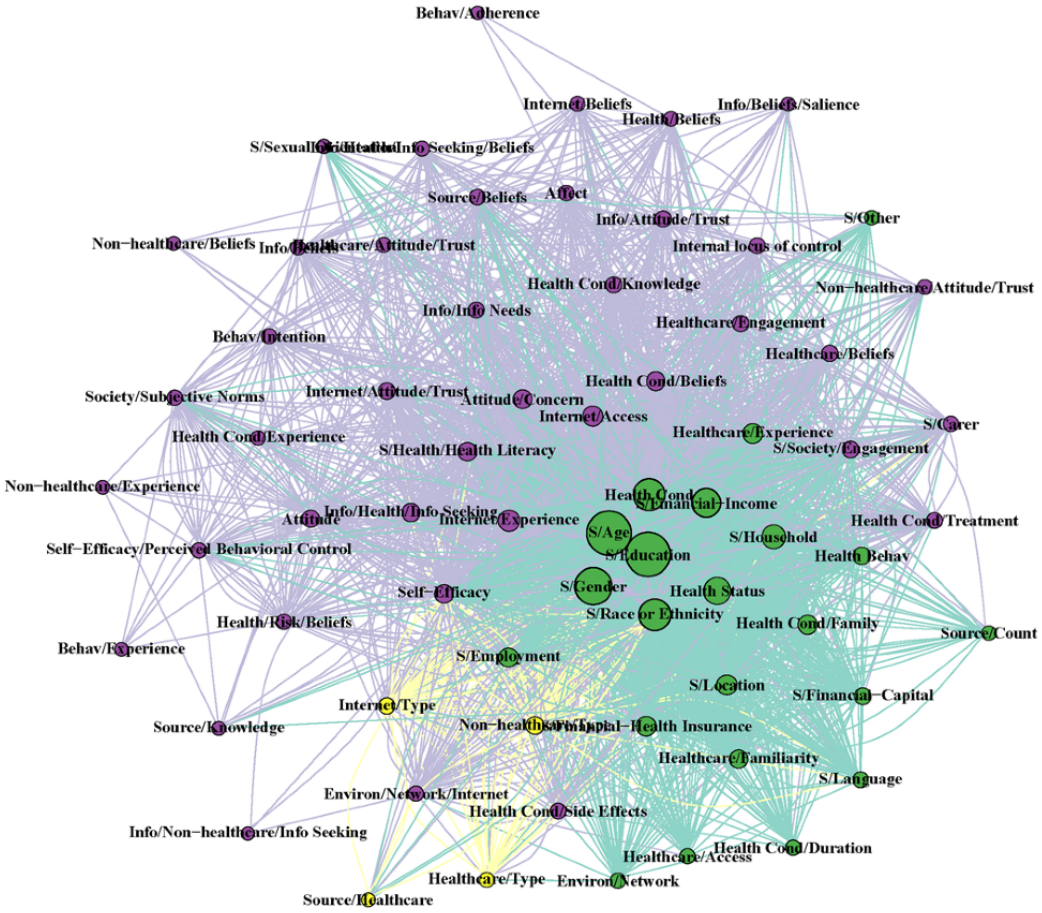
XSL•FO
RenderX

Figure 4. Fruchterman-Reingold layout algorithm of the network analysis comparing the post-2008 network model with color coding according to the modularity group membership of the complete model. The repository is available on GitHub [46].



XSL•FO
RenderX

Figure 5. Fruchterman-Reingold layout algorithm of the network analysis comparing the complete model with all years until 2019 with color coding according to the modularity group membership of the complete model. The repository is available on GitHub [46].



Discussion

Principal Findings

This paper reports a systematic literature search to identify and characterize predictors of health information–seeking behavior. The 344 included papers report 1595 significant predictor terms of health information–seeking behavior that can be classified into 68 categories. A comprehensive list of health information–seeking behavior predictor terms was developed. A novel temporal network analysis through the comparison of 2 sequential time-aggregated networks was conducted to characterize the relationships between health information–seeking behavior predictors and identify changes throughout time. This approach has never been previously used to characterize such relationships. Key findings were an increase in papers reporting on multiple significant predictors of health information–seeking behavior within a paper and a reduced rate in the identification of new predictors. The use of network analysis to map the relationships within a research field throughout time demonstrates the evolving nature of research and provides insight into how the understanding of predictors of health information–seeking behavior has developed. Van de Wijngaert et al [50] conducted a network approach to examine the current state in a research field pertaining to the adoption of eGovernment services. However, they used structural equation modeling and cross-sectional analysis, as opposed to our network comparison. The advantage of network comparison throughout time is that it allows for characterizing the evolution of research fields. Future work would be to explore the use of temporal dynamics of networks, ideally through the analysis of longitudinal data sets [43-45].

A recent meta-analysis conducted by Chang and Huang [17] on the antecedents predicting health information seeking aggregated the antecedents into 7 categories; through their review, they were able to quantify the effect sizes of the 7 categories. However, in their methodological design, not all the predictors of health information–seeking behavior were captured. Although a valuable review, we note the differences in design and the papers that could be retrieved. Since the initial review by Anker et al [16] in 2010, an increased number of papers exploring predictors of health information–seeking behavior has allowed for greater granularity in the identification of health information–seeking behavior predictors. Specifically, we have been able to develop a sociodemographic group of predictors from 5 predictors (age, gender, education, race, and health literacy) to 16 predictors (caregiver, employment, household, language, sexual orientation, finances, societal engagement, and location of residence). A potential benefit of such granularity is the improved targeting of interventions to optimize health information–seeking behavior.

Anker et al [16] reported that medication adherence is an outcome of engaging in health information seeking. However, according to our review, adherence was identified as a predictor.

XSL•FO
RenderX

This suggests that health information–seeking behavior is affected by a feedback loop, where outcomes from health information–seeking behavior can be a predictor for further health information–seeking behavior. This relationship should be examined in longitudinal studies. Few longitudinal studies have examined health information–seeking behavior, but our findings suggest that this could lead to unidentified health information–seeking behavior predictors. A longitudinal study described the reciprocal relationship between health anxiety and health information–seeking behavior on the web and how *cyberchondriac* people have health anxiety exacerbated [51]. Another study looked at clinician information engagement and information seeking [52], whereas others have shown that information needs and preferences change with time [53,54]. These initial findings demonstrate the utility of further longitudinal studies to measure additional predictors and outcomes of health information–seeking behavior.

The rate of article publication on health information–seeking behavior increased after 2005, with a doubling of articles published in the past 10 years compared with the prior 30 years. This finding mirrors the overall increase in the academic publishing rate [55]. Another explanation might be the establishment of data gathering institutions, such as the HINTS from the National Cancer Institute, which was established in 2003. Such cohort studies provide researchers with important opportunities to examine health information–seeking behaviors across large nationally representative sampling frames. HINTS is the most used data set across papers; therefore, it is the main source of identified health information–seeking behavior predictors. HINTS comprises 12 cross-sectional surveys that have been conducted in the past 15 years [56-65]. The data set has the advantage of being a representative sample frame of the United States.

The number of articles using theory to underpin their research has also increased with time. The use of theory has become a consistent theme in describing significant predictors. Interestingly, in the past 7 years, there has been a plateau in the frequency of publications reporting health information–seeking behavior predictors. A possible reason for this is maturation in the literature, with apparent saturation of identified predictors. Li et al [66] also identified an increase in publications until 2014. Our findings have extended this trend to demonstrate a plateau in publication rates since 2015.

Participants' interactions with social media, including social networking sites and health blogs, which were categorized as environment/network/internet, were identified as new predictors since 2008. Hamid et al [67] reviewed the role of social media in information-seeking behavior among international students, highlighting that specific information needs were satisfied by using social media. Although social media can be a medium for public health intervention [68], it can also present a challenge as a source of misinformation [69,70]. Competing misinformation has implications for providers of information using web media to target their audience. Providers or information creators could address the rise of misinformation by ensuring that the content delivered through social media is verified for quality and that continued monitoring is implemented.

The terminology used to describe predictors varied significantly between papers and, at times, lacked precision. Articles may have mentioned race as a predictor, but on closer inspection of the survey, items used for race, ethnicity, and culture overlap. A potential contributor to the lack of clarity in terminology is the low number of studies that used a theoretically grounded approach. Ambiguous terminology poses a challenge when comparing findings between papers on health information–seeking behavior. In response to this issue, this review developed a common classification structure for predictor terms. This structure has the potential to be developed into a future consensus taxonomy for predictors of health information–seeking behavior using domain ontologies [71].

Conducting a network analysis for predictors of health information–seeking behavior is a novel approach for analyzing the health information–seeking behavior research field. The overall network analysis shows the interrelationship of the predictor variables; however, the interaction between these variables in predicting health information–seeking behavior is still unclear. A concern is the issue of terms being correlated with each other, resulting in collinearity. The collinearity of predictor terms may affect how an individual predictor term affects the health information–seeking behavior. The temporal network analysis finding of a 6.1 greater mean degree demonstrates an increase in publications reporting multiple significant predictors. Increased reporting was also supported by an increase in co-reporting, represented by reducing network modularity with time. The high centrality of age, education, gender, and health condition indicates that these were the most commonly reported predictors when multiple predictors were reported in the included studies. Such studies have a greater ability to identify the predictors of collinearity. The network analysis approach allowed us to examine how our understanding of predictors of health information–seeking behavior and their interrelationships changed with time. Such changes have occurred in the presence of a shift toward mobile technology becoming commonplace.

Strengths and Limitations

This review has several strengths. Multiple authors followed a rigorous methodology to extract data and reach an agreement on definitions. However, because of the sheer number of articles returned in the initial database searches, there is a potential risk that articles meeting the inclusion criteria could have been potentially omitted. Nevertheless, the likelihood of omitted articles affecting the findings of this review was low because of the number of included articles. The developed taxonomy of predictors was directly informed by the included articles via a theoretical agnostic approach and consensus between 2 authors (AM and CRS). The content validity of the developed taxonomy would benefit from validation via conceptual synthesis and a consensus approach from experts in the field of health information–seeking behavior. The reliability of data extraction could be considered a limitation of our review, which we tried to mitigate through a 10% audit by a third author (CRS). Using a quantitative measure of intercoder reliability would increase confidence in reliability.

XSL•FO
RenderX

Limitations to the search strategy are, first, the inclusion of only articles published in English. This is a potential issue in this field, as geographic and cultural differences have been identified. The United States is the most represented country. However, a bibliometric analysis of the internet health information–seeking behavior literature has been previously performed by Li et al [66]. The authors similarly identified a majority of articles from the United States. A skew toward a single country may introduce geographic bias in the literature and subsequent identification of significant health information–seeking behavior predictors. There is evidence that context can directly influence individuals' health information–seeking behavior, such as being in a low-resource setting [72]. Therefore, it is important to be mindful of the number of studies from high-resource settings when considering the implications of our findings in low-resource settings. For instance, the presence of the predictor variable *health care source accessibility* may be more pertinent for countries without universal health care coverage, such as the United States, where access to physicians is variable [73,74] (Multimedia Appendix 1). A second limitation is the restriction of the definition of health information–seeking behavior as an active behavior. This limited the ability of the review findings to represent the predictors of passive health information–seeking behavior or *scanning*. Third, the review findings do not represent predictors of health information–seeking behavior for university students because of the exclusion of this subpopulation.

Finally, systematic reviews typically include an assessment of the risk of bias. The heterogeneity of the studies and the observational nature of most study designs meant that an assessment of bias was not appropriate. This led to this systematic review not adhering to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines and protocols in their entirety. However, the systematic approach adds to the strengths of this study.

Conclusions

A systematic literature search identified 344 papers reporting the predictors of health information–seeking behavior. A common taxonomy was developed to classify the predictors of health information–seeking behavior into 67 categories. Only 24% (16/67) of the predictor groupings have emerged since the invention of smartphones. Novel network analysis identified that the growth of new predictor groupings had approached saturation with only a single new predictor identified in the past 7 years, despite increasing publication rates. Publication network analysis is a promising methodology for measuring trends across scientific fields.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms, list of predictors and their definitions, network statistics, and network analysis methods.
[DOCX File , 31 KB - jmir_v23i7e21680_app1.docx]

References

1. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct 01;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
2. Lenz ER. Information seeking: a component of client decisions and health behavior. ANS Adv Nurs Sci 1984 Apr;6(3):59-72. [doi: 10.1097/00012272-198404000-00010] [Medline: 6426379]
3. Barsevick AM, Johnson JE. Preference for information and involvement, information seeking and emotional responses of women undergoing colposcopy. Res Nurs Health 1990 Feb;13(1):1-7. [doi: 10.1002/nur.4770130103] [Medline: 2305105]
4. Conley VM. Beyond knowledge deficit to a proposal for information-seeking behaviors. Nurs Diagn 1998 Apr;9(4):129-135. [doi: 10.1111/j.1744-618x.1998.tb00182.x] [Medline: 9987316]
5. Baker LM, Connor JJ. Physician-patient communication from the perspective of library and information science. Bull Med Libr Assoc 1994 Jan;82(1):36-42 [FREE Full text] [Medline: 8136759]
6. Johnson J. On contexts of information seeking. Inf Process Manag 2003 Sep;39(5):735-760. [doi: 10.1016/s0306-4573(02)00030-4]
7. Czaja R, Manfredi C, Price J. The determinants and consequences of information seeking among cancer patients. J Health Commun 2003 Nov;8(6):529-562. [doi: 10.1080/716100418] [Medline: 14690888]
8. van der Molen B. Relating information needs to the cancer experience: 1. Information as a key coping strategy. Eur J Cancer Care (Engl) 1999 Dec;8(4):238-244. [doi: 10.1046/j.1365-2354.1999.00176.x] [Medline: 10889622]
9. Rees CE, Bath PA. The psychometric properties of the Miller Behavioural Style Scale with adult daughters of women with early breast cancer: a literature review and empirical study. J Adv Nurs 2000 Aug;32(2):366-374. [doi: 10.1046/j.1365-2648.2000.01485.x] [Medline: 10964184]
10. Johnson JD, Donohue WA, Atkin CK, Johnson S. A comprehensive model of information seeking. Sci Commun 2016 Aug 18;16(3):274-303. [doi: 10.1177/1075547095016003003]
11. Cole C. Looking for information: a survey of research on information seeking, needs, and behavior (4th edition). Donald O. Case and Lisa M. Given. Bingley, UK: Emerald Group Publishing, 2016. 528 pp. $82.95 (hardcover). (ISBN: 9781785609688). J Asso Inf Sci Technol 2016 Dec 21;68(9):2284-2286. [doi: 10.1002/asi.23778]

XSL•FO
RenderX

12. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep 12;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
13. Kelly S, Eldredge SA, Dalton ED, Miller LE. Health-information behavior: an initial validity portfolio for active and passive measures. Commun Res Rep 2014 Apr 25;31(2):171-182. [doi: 10.1080/08824096.2014.907145]
14. Zare-Farashbandi F, Lalazaryan A. A review of models and theories of health information seeking behavior. Int J Health Syst Disaster Manage 2014;2(4):193. [doi: 10.4103/2347-9019.144371]
15. Simon HA. Designing organizations for an information-rich world. International Library of Critical Writings in Economics. 1969. URL: https://zeus.zeit.de/2007/39/simon.pdf [accessed 2021-06-14]
16. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. Patient Educ Couns 2011 Mar;82(3):346-354. [doi: 10.1016/j.pec.2010.12.008] [Medline: 21239134]
17. Chang C, Huang M. Antecedents predicting health information seeking: a systematic review and meta-analysis. Int J Inf Manag 2020 Oct;54:102115. [doi: 10.1016/j.ijinfomgt.2020.102115]
18. Nelissen S, Beullens K, Lemal M, Van den Bulck J. Fear of cancer is associated with cancer information seeking, scanning and avoiding: a cross-sectional study among cancer diagnosed and non-diagnosed individuals. Health Info Libr J 2015 Jun 23;32(2):107-119 [FREE Full text] [doi: 10.1111/hir.12100] [Medline: 25809822]
19. Sheldon T, Parker H. Race and ethnicity in health research. J Public Health (Oxf) 1992;14(2):104-110. [doi: 10.1093/oxfordjournals.pubmed.a042706]
20. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. Health Commun 2007 Aug 08;22(2):153-167. [doi: 10.1080/10410230701454189] [Medline: 17668995]
21. Demographics of internet and home broadband usage in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-05-28]
22. Demographics of mobile device ownership and adoption in the United States. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-05-28]
23. Anderson M. Mobile technology and home broadband 2019. Pew Research Center. 2019. URL: https://www.benton.org/headlines/mobile-technology-and-home-broadband-2019 [accessed 2021-05-28]
24. Montag C, Błaszkiewicz K, Sariyska R, Lachmann B, Andone I, Trendafilov B, et al. Smartphone usage in the 21st century: who is active on WhatsApp? BMC Res Notes 2015 Aug 04;8(1):331 [FREE Full text] [doi: 10.1186/s13104-015-1280-z] [Medline: 26238512]
25. Percheski C, Hargittai E. Health information-seeking in the digital age. J Am Coll Health 2011 Apr 13;59(5):379-386. [doi: 10.1080/07448481.2010.513406] [Medline: 21500056]
26. Huerta TR, Walker DM, Johnson T, Ford EW. A time series analysis of cancer-related information seeking: hints from the Health Information National Trends Survey (HINTS) 2003-2014. J Health Commun 2016 Sep;21(9):1031-1038. [doi: 10.1080/10810730.2016.1204381] [Medline: 27565190]
27. Rutten LJ, Agunwamba AA, Wilson P, Chawla N, Vieux S, Blanch-Hartigan D, et al. Cancer-related information seeking among cancer survivors: trends over a decade (2003-2013). J Cancer Educ 2016 Jun 26;31(2):348-357. [doi: 10.1007/s13187-015-0802-7] [Medline: 25712202]
28. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Inform J 2016 Dec 26;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
29. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2007 Sep 20;22(2):338-342. [doi: 10.1096/fj.07-9492lsf]
30. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
31. Du H, Li N, Brown MA, Peng Y, Shuai Y. A bibliographic analysis of recent solar energy literatures: the expansion and evolution of a research field. Renew Energy 2014 Jun;66:696-706. [doi: 10.1016/j.renene.2014.01.018]
32. Li J, Zhang Y, Wang X, Ho YS. Bibliometric analysis of atmospheric simulation trends in meteorology and atmospheric science journals. Croatica Chemica Acta 2009;82(3):695-705 [FREE Full text]
33. Mao N, Wang M, Ho Y. A bibliometric study of the trend in articles related to risk assessment published in science citation index. Hum Ecol Risk Assess 2010 Aug 23;16(4):801-824. [doi: 10.1080/10807039.2010.501248]
34. Wang L, Zhao L, Mao G, Zuo J, Du H. Way to accomplish low carbon development transformation: a bibliometric analysis during 1995–2014. Renew Sustain Energy Rev 2017 Feb;68:57-69. [doi: 10.1016/j.rser.2016.08.021]
35. Leppink J, Pérez-Fuster P. Social networks as an approach to systematic review. Health Prof Edu 2019 Sep;5(3):218-224 [FREE Full text] [doi: 10.1016/j.hpe.2018.09.002]
36. Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. Syst Rev 2014 Sep 29;3(1):109 [FREE Full text] [doi: 10.1186/2046-4053-3-109] [Medline: 25267336]
37. Marsden PV. Measures of network centrality. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Amsterdam: Elsevier; 2015:1-23185.
38. Cherven K. Mastering Gephi Network Visualization. Birmingham, United Kingdom: Packt Publishing; 2015:1-378.
39. Khokhar D. Gephi Cookbook. Birmingham, United Kingdom: Packt Publishing Limited; 2015.

XSL•FO
RenderX

40. Sarwar M, Soomro T. Impact of smartphone's on society. Eur J Sci Res 2013;98(2):216-226 [FREE Full text]
41. Islam N, Want R. Smartphones: past, present, and future. IEEE Pervasive Comput 2014 Oct;13(4):89-92. [doi: 10.1109/mprv.2014.74]
42. The World in 2009: ICT facts and figures. ITU-D. 2009. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2009.pdf [accessed 2021-05-28]
43. Blonder B, Wey TW, Dornhaus A, James R, Sih A. Temporal dynamics and network analysis. Methods Ecol Evol 2012 Aug 01;3(6):958-972. [doi: 10.1111/j.2041-210x.2012.00236.x]
44. Santoro N, Quattrociocchi W, Flocchini P, Casteigts A, Amblard F. Time-varying graphs and social network analysis: temporal indicators and metrics. arXiv : Computer Science - Social and Information Networks. 2011. URL: https://arxiv.org/abs/1102.0629 [accessed 2021-05-28]
45. Tang J, Musolesi M, Mascolo C, Latora V. Temporal distance metrics for social network analysis. In: Proceedings of the 2nd ACM Workshop on Online Social Networks. 2009 Presented at: SIGCOMM '09: ACM SIGCOMM 2009 Conference; August, 2009; Barcelona Spain p. 31-36. [doi: 10.1145/1592665.1592674]
46. Mirzaei A. Algorithm and code used for calculating and displaying a network analysis for the predictors of health information seeking behaviour. Github: HISB Predictors Network Analysis. 2020. URL: https://github.com/ardimirzaei/HISB_Predictors_Network_Analysis [accessed 2021-05-28]
47. Csardi G, Nepusz T. The igraph software package for complex network research. InterJ Compl Sys 2006;1695(5):1-9 [FREE Full text]
48. Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp 1991 Nov;21(11) [FREE Full text]
49. Mullarkey MC, Marchetti I, Beevers CG. Using network analysis to identify central symptoms of adolescent depression. J Clin Child Adolesc Psychol 2019;48(4):656-668 [FREE Full text] [doi: 10.1080/15374416.2018.1437735] [Medline: 29533089]
50. van de Wijngaert L, Bouwman H, Contractor N. A network approach toward literature review. Qual Quant 2012 Nov 10;48(2):623-643. [doi: 10.1007/s11135-012-9791-3]
51. Poel FT, Baumgartner SE, Hartmann T, Tanis M. The curious case of cyberchondria: a longitudinal study on the reciprocal relationship between health anxiety and online health information seeking. J Anxiety Disord 2016 Oct;43:32-40. [doi: 10.1016/j.janxdis.2016.07.009] [Medline: 27497667]
52. Moldovan-Johnson M, Tan AS, Hornik RC. Navigating the cancer information environment: the reciprocal relationship between patient-clinician information engagement and information seeking from nonmedical sources. Health Commun 2014 Dec 20;29(10):974-983 [FREE Full text] [doi: 10.1080/10410236.2013.822770] [Medline: 24359259]
53. Harrison DE, Galloway S, Graydon JE, Palmer-Wickham S, Rich-van der Bij L. Information needs and preference for information of women with breast cancer over a first course of radiation therapy. Patient Educ Couns 1999 Nov;38(3):217-225. [doi: 10.1016/s0738-3991(99)00009-9]
54. Matsuyama RK, Kuhn LA, Molisani A, Wilson-Genderson MC. Cancer patients' information needs the first nine months after diagnosis. Patient Educ Couns 2013 Jan;90(1):96-102. [doi: 10.1016/j.pec.2012.09.009] [Medline: 23058682]
55. Bjork B, Roos A, Lauri M. Scientific journal publishing: yearly volume and open access availability. Information Research. 2009. URL: http://www.informationr.net/ir/14-1/paper391.html [accessed 2021-05-28]
56. National Cancer Institute. Health Information National Trends Survey. NCI Thesaurus, Code C20093 2003:A [FREE Full text] [doi: 10.32388/6P9GLV]
57. Westat. Health Information National Trends Survey 2005 (HINTS 2005) - final report. National Cancer Institute, Bethesda, MD. 2005. URL: https://hints.cancer.gov/docs/HINTS_2005_Final_Report.pdf [accessed 2021-05-28]
58. Westat. Health Information National Trends Survey 4 (HINTS 4) Cycle 1 Methodology Report. National Cancer Institute, Bethesda, MD. 2012. URL: https://hints.cancer.gov/docs/HINTS4_Cycle1_Methods_Report_revised_Jun2012.pdf [accessed 2021-05-28]
59. Westat. Health Information National Trends Survey 4 (HINTS 4) Cycle 2 Methodology Report. National Cancer Institute, Bethesda, MD. 2013. URL: https://hints.cancer.gov/docs/HINTS_4_Cycle2_Methods_Report.pdf [accessed 2021-05-28]
60. Westat. Health Information National Trends Survey 4 (HINTS 4) Cycle 3 Methodology Report. National Cancer Institute, Bethesda, MD. 2014. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS_4_Cycle_3_Methodology_Report.pdf [accessed 2021-05-28]
61. Westat. Health Information National Trends Survey 4 (HINTS 4) HINTS-FDA Methodology Report. National Cancer Institute Bethesda, MD. 2015. URL: https://hints.cancer.gov/docs/Instruments/HINTS-FDA_Methodology_Report.pdf [accessed 2021-05-28]
62. Westat. Health Information National Trends Survey 4 (HINTS 4) HINTS-FDA2 Methodology Report. National Cancer Institute Bethesda, MD. 2017. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS_FDA_Cycle2_Methodology_Rpt.pdf [accessed 2021-05-28]
63. Westat. Health Information National Trends Survey 5 (HINTS 5) Cycle 1 Methodology Report. National Cancer Institute Bethesda, MD. 2017. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle_1_Methodology_Rpt.pdf [accessed 2021-05-28]

XSL•FO
RenderX

64. Westat. Health Information National Trends Survey 5 (HINTS 5) Cycle 2 Methodology Report. National Cancer Institute Bethesda, MD. 2018. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle_2_Methodology_Report.pdf [accessed 2021-05-28]
65. Westat. Health Information National Trends Survey 4 (HINTS 4) Cycle 4 Methodology Report. National Cancer Institute Bethesda, MD. 2015. URL: https://hints.cancer.gov/docs/Instruments/HINTS_4_Cycle_4_Methodology_Report_Final.pdf [accessed 2021-05-28]
66. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
67. Hamid S, Bukhari S, Ravana SD, Norman AA, Ijab MT. Role of social media in information-seeking behaviour of international students. Aslib J Inf Manag 2016 Sep 19;68(5):643-666. [doi: 10.1108/ajim-03-2016-0031]
68. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018 Sep 19;1(1):1-47 [FREE Full text] [doi: 10.1038/s41746-018-0054-0] [Medline: 30854472]
69. Rolls K, Massey D. Social media is a source of health-related misinformation. Evid Based Nurs 2021 Apr 11;24(2):46. [doi: 10.1136/ebnurs-2019-103222] [Medline: 32046968]
70. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
71. McDaniel M, Storey VC. Evaluating domain ontologies. ACM Comput Surv 2019 Sep 18;52(4):1-44. [doi: 10.1145/3329124]
72. Dutta R. Information needs and information-seeking behavior in developing countries: a review of the research. Int Inf Library Rev 2013 Dec 02;41(1):44-51. [doi: 10.1080/10572317.2009.10762796]
73. Naylor KB, Tootoo J, Yakusheva O, Shipman SA, Bynum JP, Davis MA. Geographic variation in spatial accessibility of U.S. healthcare providers. PLoS One 2019 Apr 9;14(4):e0215016 [FREE Full text] [doi: 10.1371/journal.pone.0215016] [Medline: 30964933]
74. Aldridge ML, Daniels JL, Jukic AM. Mammograms and healthcare access among US Hispanic and non-Hispanic women 40 years and older. Fam Community Health 2006;29(2):80-88. [doi: 10.1097/00003727-200604000-00004] [Medline: 16552286]

Abbreviations

HINTS: Health Information Trends Survey
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by R Kukafka, G Eysenbach; submitted 21.06.20; peer-reviewed by CA Almenara, R Haase, W Pian, PCI Pang, YH Liu, Y Cheng; comments to author 27.10.20; revised version received 30.11.20; accepted 12.05.21; published 02.07.21.

Please cite as:
Mirzaei A, Aslani P, Luca EJ, Schneider CR
Predictors of Health Information–Seeking Behavior: Systematic Literature Review and Network Analysis
J Med Internet Res 2021;23(7):e21680
URL: https://www.jmir.org/2021/7/e21680
doi:10.2196/21680
PMID:33979776

©Ardalan Mirzaei, Parisa Aslani, Edward Joseph Luca, Carl Richard Schneider. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis

Jo-Hsuan Wu[1], MD; T Y Alvin Liu[2], MD; Wan-Ting Hsu[3], MSc; Jennifer Hui-Chun Ho[4], PhD, MD; Chien-Chang Lee[5,6,7], SCD, MD

[1]Shiley Eye Institute and Viterbi Family Department of Ophthalmology, University of California San Diego, La Jolla, CA, United States

[2]Retina Division, Wilmer Eye Institute, The Johns Hopkins Medicine, Baltimore, MD, United States

[3]Harvard TH Chan School of Public Health, Boston, MA, United States

[4]National Yang-Ming University, Taipei City, Taiwan

[5]Health Data Science Research Group, National Taiwan University Hospital, Taipei, Taiwan

[6]The Centre for Intelligent Healthcare, National Taiwan University Hospital, Taipei, Taiwan

[7]Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan

Corresponding Author:
Chien-Chang Lee, SCD, MD
Department of Emergency Medicine, National Taiwan University Hospital
No 7, Chung-Shan South Road
Taipei, 100
Taiwan
Phone: 886 2 23123456 ext 63485
Fax: 886 2 23223150
Email: hit3transparency@gmail.com

Abstract

Background: Diabetic retinopathy (DR), whose standard diagnosis is performed by human experts, has high prevalence and requires a more efficient screening method. Although machine learning (ML)–based automated DR diagnosis has gained attention due to recent approval of IDx-DR, performance of this tool has not been examined systematically, and the best ML technique for use in a real-world setting has not been discussed.

Objective: The aim of this study was to systematically examine the overall diagnostic accuracy of ML in diagnosing DR of different categories based on color fundus photographs and to determine the state-of-the-art ML approach.

Methods: Published studies in PubMed and EMBASE were searched from inception to June 2020. Studies were screened for relevant outcomes, publication types, and data sufficiency, and a total of 60 out of 2128 (2.82%) studies were retrieved after study selection. Extraction of data was performed by 2 authors according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), and the quality assessment was performed according to the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2). Meta-analysis of diagnostic accuracy was pooled using a bivariate random effects model. The main outcomes included diagnostic accuracy, sensitivity, and specificity of ML in diagnosing DR based on color fundus photographs, as well as the performances of different major types of ML algorithms.

Results: The primary meta-analysis included 60 color fundus photograph studies (445,175 interpretations). Overall, ML demonstrated high accuracy in diagnosing DR of various categories, with a pooled area under the receiver operating characteristic (AUROC) ranging from 0.97 (95% CI 0.96-0.99) to 0.99 (95% CI 0.98-1.00). The performance of ML in detecting more-than-mild DR was robust (sensitivity 0.95; AUROC 0.97), and by subgroup analyses, we observed that robust performance of ML was not limited to benchmark data sets (sensitivity 0.92; AUROC 0.96) but could be generalized to images collected in clinical practice (sensitivity 0.97; AUROC 0.97). Neural network was the most widely used method, and the subgroup analysis revealed a pooled AUROC of 0.98 (95% CI 0.96-0.99) for studies that used neural networks to diagnose more-than-mild DR.

Conclusions: This meta-analysis demonstrated high diagnostic accuracy of ML algorithms in detecting DR on color fundus photographs, suggesting that state-of-the-art, ML-based DR screening algorithms are likely ready for clinical applications. However, a significant portion of the earlier published studies had methodology flaws, such as the lack of external validation and presence of spectrum bias. The results of these studies should be interpreted with caution.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e23863) doi:10.2196/23863

KEYWORDS

machine learning; diabetic retinopathy; diabetes; deep learning; neural network; diagnostic accuracy

Introduction

Diabetic retinopathy (DR) is the leading cause of vision impairment and blindness among working-aged people in the world [1]. Approximately one-third of people with diabetes mellitus have signs of DR, among whom one-third have vision-threatening DR (VTDR). A meta-analysis estimated global prevalence of any DR and proliferative diabetic retinopathy (PDR) among patients with diabetes to be 35.4% and 7.5%, respectively [2]. The number of patients with DR is approximately 93 million and is expected to rise to 191 million by 2030, as type 2 diabetes has attained the status of a global pandemic, spreading from affluent industrialized nations to the developing world [3].

Vision impairment due to DR can be significantly reduced if diagnosed in early stages and treated appropriately [4]. However, fewer than 60% of patients with diabetes undergo regular eye examinations at intervals recommended by guidelines due to the high cost and low accessibility of ophthalmologic services [3]. The number of people with diabetes that need regular eye examinations has quadrupled in the past three decades. Therefore, the development of an automatic, low-cost, accurate eye screening tool has become an important public health issue [5]. The gold standard for DR screening is based on clinical examinations by human clinicians or the analysis of color fundus photographs via telemedicine [6]. However, both approaches are time-consuming, labor-intensive, and prone to inconsistency due to inherent human subjectivity [7]. Automated systems that are capable of interpreting color fundus photographs with high sensitivity and specificity are critical for widespread implementation of DR screening, and the rise of artificial intelligence (AI), specifically machine learning (ML), has made such automated approaches a possibility.

ML uses existing data to train a computer to recognize a specific pattern or predict a specific outcome in a new data set [6]. Exploration of automated image analysis can be dated back to 1980, when classical ML methods, such as support vector machines and random forests, were used to detect predefined features [8]. These early ML techniques for detecting DR employed mathematical image transformation techniques and image engineering guided by expert-designed rules [9]. The accuracy of this type of analysis did not reach the standard of clinical application. In recent years, the advent of deep learning (DL), a subtype of ML, has transformed the field of automated image analysis [10]. Briefly, DL methods are representation learning methods that use multilayered neural networks, the performance of which can be enhanced by reiteratively changing the internal parameters [11,12]. Unlike other ML approaches, DL does not require image engineering. It develops its own representations needed for pattern recognition after being fed raw data and has shown superior accuracy as compared with other classical ML algorithms [13,14].

Although ML has garnered significant attention with the recent US Food and Drug Administration (FDA) approval of the first ML-based, fully automatic DR screening machine in April 2018 [15], skepticism within the medical community remains regarding the robustness of ML techniques in real-world clinical applications. Given that ophthalmology is among the medical disciplines that have reaped the most benefits from recent AI advancements and that DR screening is one of the most promising ML applications in ophthalmology, we have set out to systematically survey, through meta-analysis, the current status of ML as applied in DR screening based on color fundus photographs. Specifically, we have examined the range of performances reported by different studies and have determined which ML technique is superior for this clinical purpose.

Methods

Search Methods for Identifying Studies

This meta-analysis was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [16]. A literature search for relevant studies published through June 2020 was performed with 2 publicly available databases, PubMed and EMBASE. There were 3 stages to the literature search. No language or population filters were applied, while nonhuman experiments, case reports, guidelines, conference papers, letters, editorials, and review articles were excluded. Filter for publication year was applied only in the second and third stages of the literature search in order to avoid overlapping of search results. Duplicated references in different stages of the literature search were manually excluded. The major search key combination terms were “diabetic retinopathy” OR “diabetic macular edema” OR “macular edema” OR “retinopathy” OR “neovascularized retinopathy” OR “proliferative retinopathy” OR “referable diabetic retinopathy” OR “diabetic macular oedema” OR “proliferative diabetic retinopathy” OR “retinal disorders” OR “diabetic eye disease” OR “vision loss” OR “retinal diseases” OR “macular disease” OR “macular degeneration” OR “macular disorders” crossed with “artificial intelligence” OR “deep learning” OR “transfer learning” OR “machine learning” OR “deep learning system”. The detailed search strategy is provided in Multimedia Appendix 1.

Eligibility Criteria for Considering Studies for This Review

We included studies that evaluated ML algorithms on the accuracy of automated image analysis for screening or diagnosis of DR. We included studies that detected pathological findings of DR, diagnosed DR status, and staged DR severity.

Study Selection

The study selection and data extraction were independently performed by 2 authors (JHW and CCL). After duplicates were removed, titles and abstracts were screened for exclusion of

XSL•FO
RenderX

studies with potentially nonrelevant outcome or publication types and studies applying information other than images in analytical work. When there were multiples studies derived from the same cohort with overlapping study periods, earlier studies were considered duplicates and only the study with the most recent result was included. Retrieved studies with accessible full articles then underwent full-text review. Discrepancies between the reviewers were resolved first by a consensus meeting and then arbitration by a third reviewer if consensus could not be reached.

Data Collection

Data extraction was performed on studies selected after full-text review. A thorough review of each article was performed with the following variables extracted: first author, published year, country, algorithm, image modality, total image size, relevant image size, number of participants and eyes, number of diseased participants and eyes, databases and characteristics, and the sensitivity and specificity of both training and validation sessions. When multiple algorithms were tested on one data set, only the data of the best-performing algorithm were included. Algorithms applied were further classified into 4 main categories: support vector machine (SVM), neural network (NN), random forest (RF), and others. After data extraction, studies were classified into different outcomes of DR, including any DR, more-than-mild diabetic retinopathy (mtmDR), vision-threatening diabetic retinopathy (VTDR), diabetic macular edema, and proliferative diabetic retinopathy (PDR). Studies with relevant data were examined for sufficiency to construct a 2 × 2 contingency table before quality assessment.

Risk of Bias Assessment

The quality of eligible studies was independently assessed by 2 reviewers using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool, which is composed of 4 domains assessing both risk of bias and applicability of clinical practice: patient selection, index test, reference standard, and flow and timing. For each diagnostic study, we determined the risk for bias and general applicability in all 4 domains of QUADAS-2 and reported them separately. A study was considered to have a low risk of bias in one domain if at least half of the variables extracted from the validation session met the requirements of QUADAS-2.

Data Synthesis and Analysis

Meta-analysis for diagnostic accuracy of each ML algorithm or DR outcome was performed with a bivariate random effects model to account for both within- and between-study heterogeneity. Results were summarized by using hierarchical summary receiver operating characteristic (ROC) plots and coupled forest plots. Pooled sensitivity, specificity, area under the curve, and positive and negative likelihood ratios were calculated. The bivariate model approach modeled the logit-transformed sensitivity and specificity simultaneously to account for the inherent negative correlation between sensitivity and specificity that might have arisen due to different thresholds in different studies. Heterogeneity was tested using the Cochran Q statistic (P<.05) and quantified with the I^2 statistic, which describes the variation of effect size that is attributable to heterogeneity across studies. For direct clinical interpretation, we also calculated the posttest probability for each type of lesion. We used the prevalence of lesions in the pooled study population as the informative prior, and derived the posttest probability of each type of lesion based on the pooled positive and negative likelihood ratios. Results are presented as Fagan nomograms. The presence and effect of publication bias were examined with Deeks tests. When publication bias is present, Deeks funnel plots are usually asymmetrical. We used a trim-and-fill method to impute hypothetical missing studies due to publication bias. Trim-and-fill odds ratios (ORs) were reported when the tests for publication bias were significant. We performed sensitivity analyses to examine the potential effects of different demographic factors, ML algorithms, and types of training or validating databases. All analyses except for the summary ROC curve were conducted by the “mada” package in R software (The R Foundation for Statistical Computing). Summary ROC and area under the ROC (AUROC) were calculated by the “midas” package in Stata 14.0 (StataCorp). A 2-sided P value <0.05 indicated statistical significance for all tests.

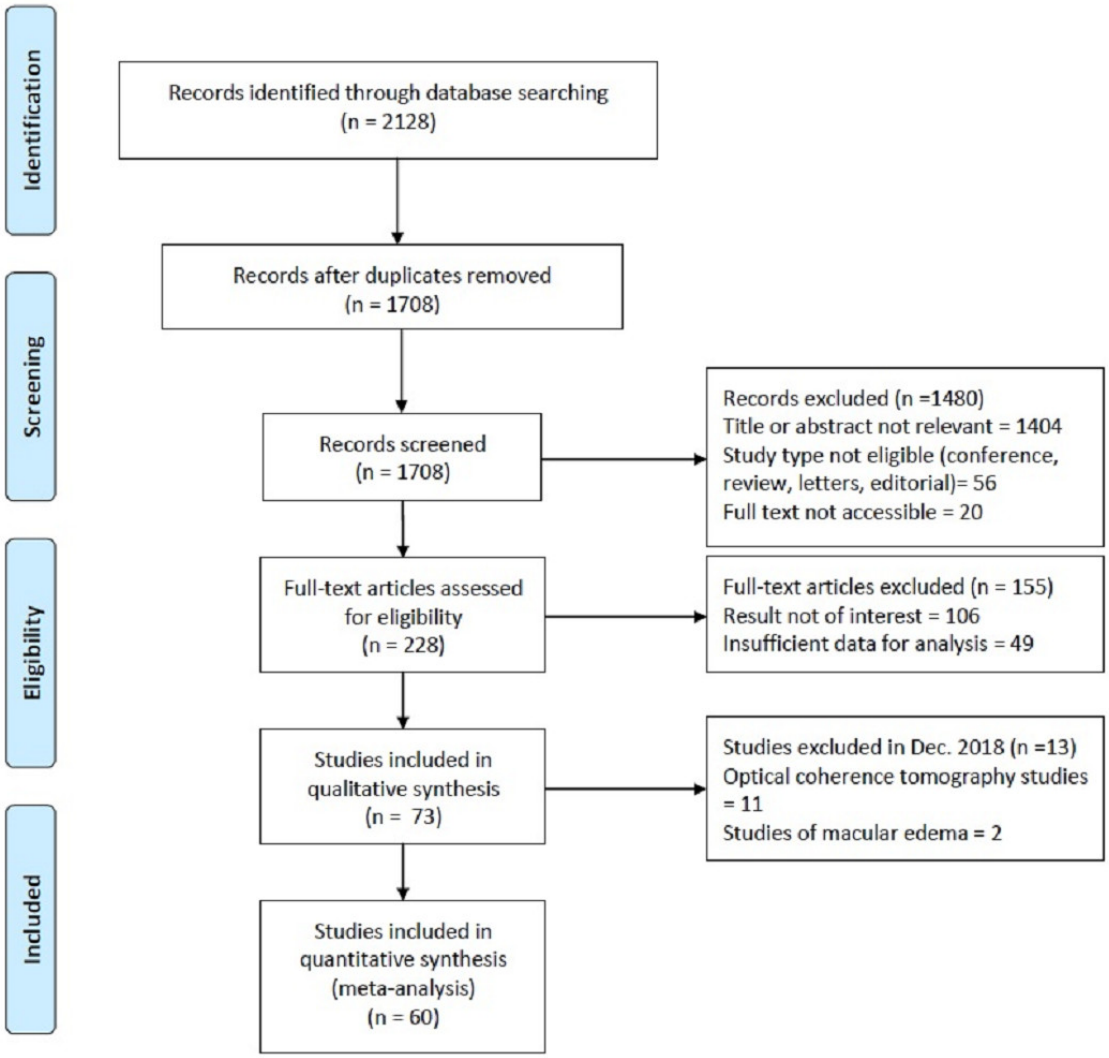
Results

Search Results

The first 2 stages of the literature search (performed in May 2018 and December 2018) yielded 668 hits from PubMed and 809 hits from EMBASE. After screening titles and abstracts, we excluded 336 studies for duplication and 941 studies for nonrelevant abstracts or publication types. A total of 187 studies went through full-text review, 99 of which were excluded for a result that was not of interest and 43 of which were excluded for insufficient data for analysis. After completing qualitative synthesis of the 45 studies, we proceeded to only include ML studies that involved use of color fundus photograph for DR screening. After further exclusion, only 32 studies were retrieved after the second stage of literature selection [13,15,17-46]. The third stage of the literature search was performed in June 2020, and 28 out of 651 studies were retrieved after literature selection [47-74], resulting in a total of 60 (N=32+28) included studies for final analysis. A new category composed of VTDR and PDR (VTDR+PDR) was created for examination of diagnostic accuracy of the most treatment-urgent group. A flowchart of the literature search and study selection process is summarized in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart of study selection.



Study Characteristics

Multimedia Appendix 2 Table S1 summarizes the study-level characteristics of studies assessing the diagnostic accuracy of ML algorithms for different categories of DR. Of the 60 studies, 35 studies (58%) evaluated any DR, 23 (38%) mtmDR, 12 (20%) VTDR, and 12 (20%) PDR. Publicly available benchmark databases, such as Messidor, Structured Analysis of the Retina (STARE), Digital Retinal Images for Vessel Extraction (DRIVE), DIARETDB, e-Ophtha, and EyePACs were used for testing of the ML algorithms in 40 of the 60 (67%) studies. The characteristics of these publicly available retinal image databases are summarized in Multimedia Appendix 3. The distribution of categories of ML algorithms used was as follows: SVM (6/60, 10%), RF (2/60, 3%), NN (37/60, 62%), and others (17/60, 28%). The general principles of these ML algorithms are described in Multimedia Appendix 4.

Quality Assessment

Quality assessments using the QUADAS-2 criteria are summarized in Multimedia Appendix 5. Most studies (56/60, 93%) presented a clear source of patient recruitment or selection criteria and processes, and were at a low risk for bias. Of the 60 studies, 3 (5%) reported limited information on the establishment of reference standard and were at a high risk for bias, and 4 (7%) reported insufficient blinding to a reference standard during interpretation of the index test results and were at a high risk for bias. For study applicability, 1 study (2%) in the index test section and 4 (7%) studies in the patient selection section were recorded to be at a high risk of concern, due to insufficient information reported.

Synthesis of Results

A summary of data of included studies is presented in Multimedia Appendix 6. Pooled sensitivities, specificities, likelihood ratios, AUROCs, and I^2 statistics for the 5 DR categories, including any DR, mtmDR, VTDR, PDR, and VTDR+PDR, are presented in Table 1. As some studies might have used more than 2 data sets for validation, performance of ML derived from each data set was viewed as individual data, and we used “data” as the unit for calculation (eg, 35 included studies performed evaluation of ML on identifying any DR, resulting in a total of 53 data for synthesis and analysis). The

XSL•FO
RenderX

hierarchical summary ROC plots for the 4 main DR categories, including DR, mtmDR, VTDR, and PDR, are also presented (Figures 2-5). ML showed a high overall accuracy in detecting the 5 categories of DR, with a pooled AUROC ranging from 0.97 (95% CI 0.96-0.99) for mtmDR and VTDR+PDR to 0.99 (95% CI 0.98-1.00) for VTDR and PDR. The pooled sensitivity for all 5 categories was high, ranging from 0.93 (95% CI 0.87-0.96) for PDR to 0.97 (95% CI 0.94-0.99) for VTDR. The pooled specificity, however, showed more variation: from 0.90 (95% CI 0.87-0.93) for mtmDR to 0.98 (95% CI 0.96-0.99) for PDR. The Fagan plots for different DR categories are presented in Multimedia Appendix 7. For images that were classified as positive by the ML algorithms, the posttest probability for DR, mtmDR, VTDR, and PDR was 87%, 71%, 66%, and 77%, respectively. For images that were classified as negative by the ML algorithms, the posttest probability for DR, mtmDR, VTDR, and PDR was 4%, 1%, 0%, and 1%, respectively.

Table 1. Pooled analysis for diagnostic accuracy of diabetic retinopathy by machine learning on color fundus photographs.

Goal of detection	Data[a], n	Sen[b,c]	Spe[d,c]	LR+[e,c]	LR–[f,c]	AUROC[g,c]	I^2 statistic[c]	Publication bias (*P* value)
Any DR[h]	53	0.94 (0.91-0.96)	0.92 (0.88-0.95)	12.4 (8.0-19.3)	0.07 (0.05-0.09)	0.98 (0.96-0.99)	32 (22-42)	.01
mtmDR[i]	40	0.95 (0.93-0.97)	0.90 (0.87-0.93)	9.7 (7.4-12.7)	0.05 (0.04-0.08)	0.97 (0.96-0.99)	29 (18-40)	.11
VTDR[j]	15	0.97 (0.94-0.99)	0.94 (0.87-0.98)	17.3 (7.5-39.9)	0.03 (0.01-0.06)	0.99 (0.98-1.00)	32 (9-56)	.33
PDR[k]	22	0.93 (0.87-0.96)	0.98 (0.96-0.99)	38.5 (21.7-68.4)	0.07 (0.04-0.13)	0.99 (0.98-1.00)	29 (11-46)	.11
VTDR and PDR	37	0.96 (0.93-0.98)	0.97 (0.94-0.98)	24.3 (14.5-38.5)	0.07 (0.05-0.10)	0.97 (0.96-0.99)	N/A	.06

[a]Machine learning data derived from each data set was viewed as individual data, and we used “data” as the unit for calculation.

[b]Sen: sensitivity.

[c]Values in this column are as follows: mean (95% confidence interval).

[d]Spe: specificity.

[e]LR+: positive likelihood ratio.

[f]LR–: negative likelihood ratio.

[g]AUROC: area under the receiver operating characteristic.

[h]DR: diabetic retinopathy.

[i]mtmDR: more-than-mild diabetic retinopathy.

[j]VTDR: vision-threatening diabetic retinopathy.

[k]PDR: proliferative diabetic retinopathy.

XSL•FO
RenderX

Figure 2. SROC curves for diagnosis of any diabetic retinopathy on color fundus photographs. AUC: area under the curve; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristics.

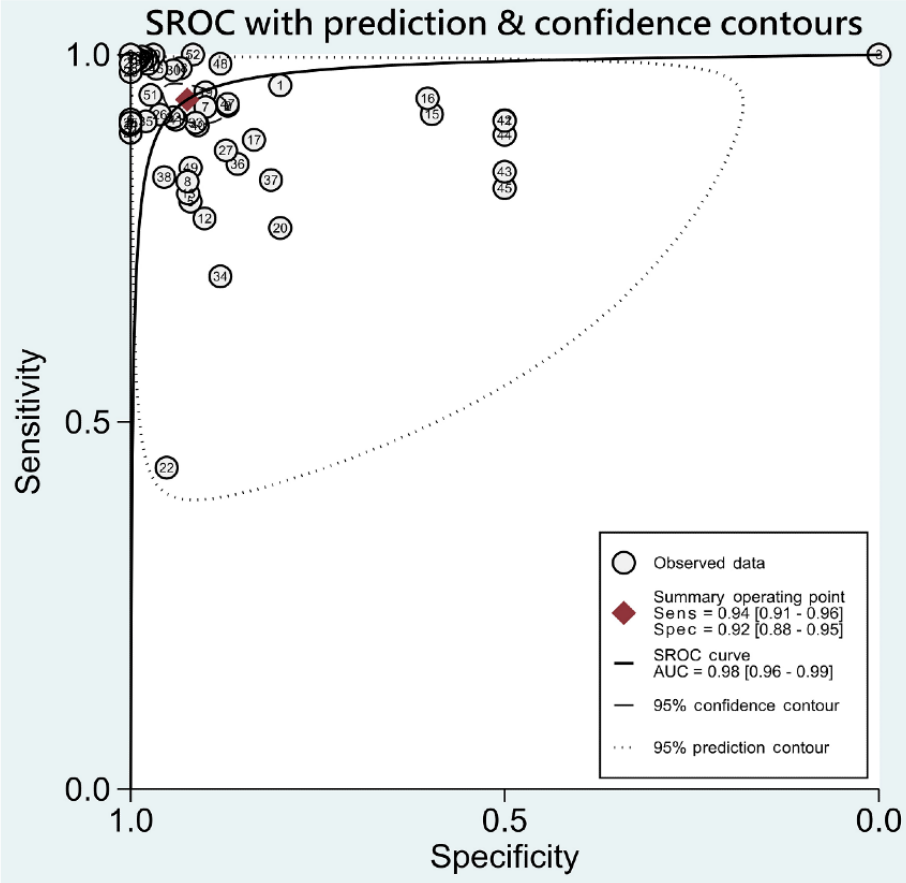
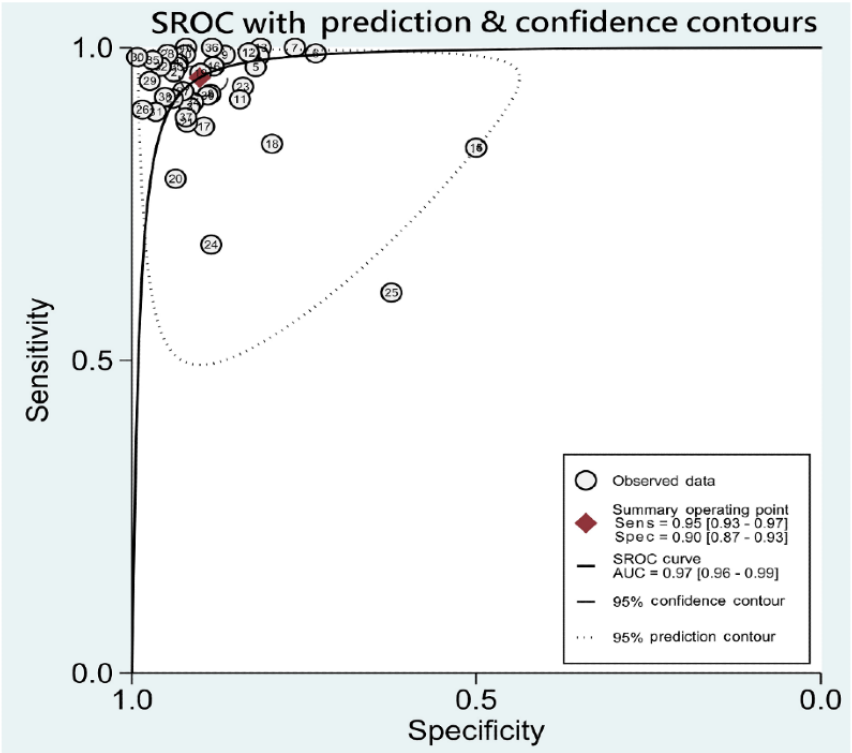


Figure 3. SROC curves for diagnosis of more-than-mild diabetic retinopathy on color fundus photographs. Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristics; AUC: area under the curve.



XSL•FO
RenderX

Figure 4. SROC curves for diagnosis of vision-threatening diabetic retinopathy on color fundus photographs. AUC: area under the curve; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristics.

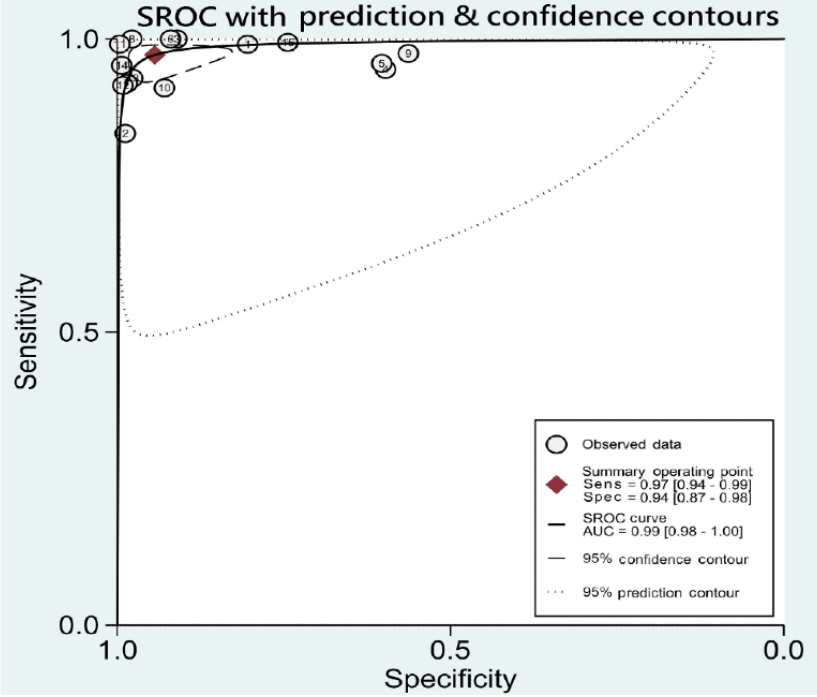
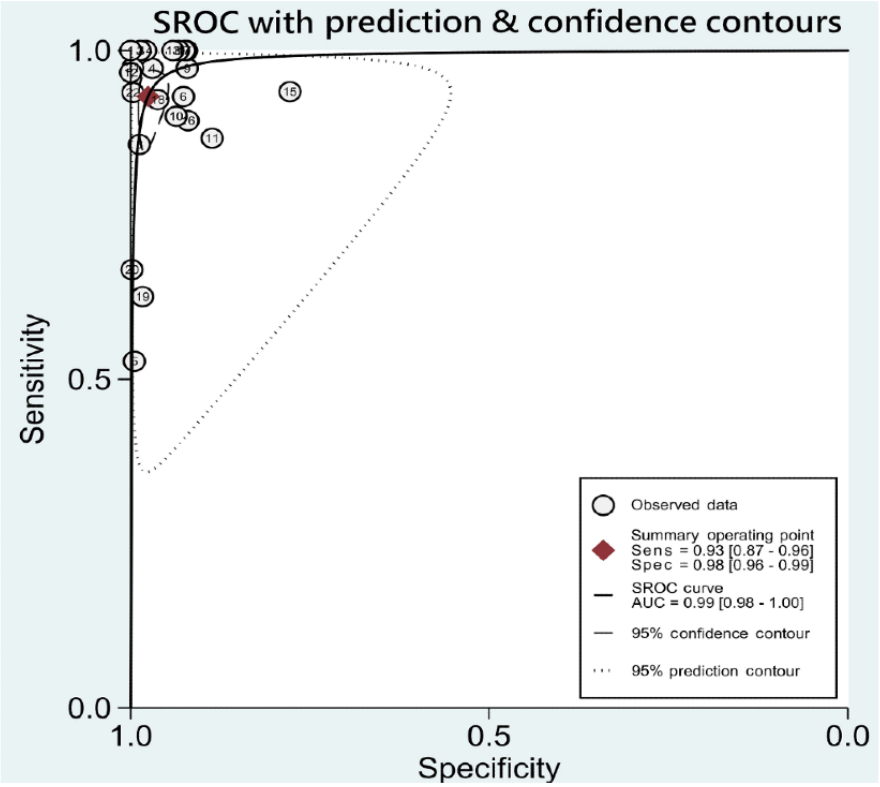


Figure 5. SROC curves for diagnosis of proliferative diabetic retinopathy on color fundus photographs. AUC: area under the curve; Sens: sensitivity; Spec: specificity; SROC: summary receiver operating characteristics.



Subgroup Analyses and Sensitivity Analysis

We performed subgroup analyses for mtmDR studies to explore the possible heterogeneity in test accuracy (Table 2). The main causes of heterogeneity included in the analysis were algorithm type, mean age of subject populations, and validation set selection. For this subgroup analysis, 23 studies were included, with a total of 40 data obtained from different testing data sets. Of the 23 studies, the 22 studies that applied NN algorithms demonstrated high pooled performance (summary AUROC 0.98; 95% CI 0.96-0.99), sensitivity (sensitivity 0.95; 95% CI 0.93-0.97), and specificity (specificity 0.91; 95% CI 0.88-0.93). The only study that used a different kind of ML algorithm (instance learning) reported significantly inferior sensitivity

XSL•FO
RenderX

(0.84) under preset specificity (0.50). Of the 60 studies, 19 (83%) tested the algorithm's performance on data sets with subject populations with a mean age greater than 50 years. Pooled sensitivity of data from these studies was high (sensitivity 0.95; 95% CI 0.93-0.97), and the pooled specificity was moderate (specificity 0.89; 95% CI 0.85-0.92). Compared with algorithms that used benchmark data sets for validation (pooled sensitivity 0.92; 95% CI 0.87-0.95), the pooled sensitivities of algorithms validated by clinical data sets (sensitivity 0.97; 95% CI 0.95-0.98) and independent data sets (sensitivity 0.96; 95% CI 0.93-0.97) were not inferior. The results of pooled AUROCs validated by these 3 types of data sets were similar, implying that robust performance of ML algorithms can be generalized to images collected in clinical practice.

Table 2. Subgroup analysis for diagnostic accuracy of mtmDR retinopathy on color fundus photographs.

Features of subgroup	Data[a], n	Sen[b,c]	Spe[d,c]	LR+[e,c]	LR–[f,c]	AUROC[g,c]	I^2 statistic[c]	Publication bias (*P* value)
Overall mtmDR[h]	40	0.95 (0.93, 0.97)	0.90 (0.87, 0.93)	9.7 (7.4, 12.7)	0.05 (0.04, 0.08)	0.97 (0.96, 0.99)	29 (18, 40)	.11
Mean age> 50 years	32	0.95 (0.93, 0.97)	0.89 (0.85, 0.92)	8.8 (6.4, 12.0)	0.05 (0.03, 0.08)	0.97 (0.95, 0.98)	33 (20, 46)	.22
NN[i] algorithms	38	0.95 (0.93, 0.97)	0.91 (0.88, 0.93)	10.1 (7.7, 13.2)	0.05 (0.03, 0.07)	0.98 (0.96, 0.99)	30 (19, 41)	.14
Benchmark test sets	15	0.92 (0.87, 0.95)	0.90 (0.82, 0.94)	9.0 (4.8, 16.6)	0.09 (0.05, 0.16)	0.96 (0.94, 0.98)	25 (10, 39)	.22
Clinical Test sets	25	0.97 (0.95, 0.98)	0.90 (0.88, 0.92)	10.0 (7.9, 12.6)	0.04 (0.02, 0.06)	0.97 (0.96, 0.98)	30 (15, 45)	.06
External validation	31	0.96 (0.93, 0.97)	0.90 (0.87, 0.93)	9.7 (7.2, 13.0)	0.05 (0.03, 0.07)	0.98 (0.96, 0.99)	29 (16, 42)	.08

[a]Machine learning data derived from each data set was viewed as individual data, and we used "data" as the unit for calculation.

[b]Sen: sensitivity.

[c]Values in this column are as follows: mean (95% confidence interval).

[d]Spe: specificity.

[e]LR+: positive likelihood ratio.

[f]LR–: negative likelihood ratio.

[g]AUROC: area under the receiver operating characteristic.

[h]mtmDR: more-than-mild diabetic retinopathy.

[i]NN: neural network.

Publication Bias

The test for publication bias was generally not significant in different categories of DR (Deeks test *P*=.01; Multimedia Appendix 8), except for any DR. Trim-and-fill analysis showed the diagnostic OR remained insignificant (OR 0.50, 95% CI 0.25-1.01) after hypothetical unpublished data were included for analysis (Multimedia Appendix 9).

Discussion

Principal Results and Comparison With Prior Work

This systematic review synthesizes the available evidence and compares the diagnostic accuracy of ML algorithms for the detection of DR based on color fundus photographs. The primary meta-analysis included 60 studies with 445,175 interpretations. Out of the 60 studies, 35 (58%) were validated by external testing data sets that were completely independent of the training data sets. Overall, ML demonstrated a robust performance in detecting different DR categories, with a pooled sensitivity of 0.93 to 0.97 and a pooled specificity of 0.90 to 0.98. The pooled sensitivity compares favorably to reported sensitivities of 73% [75], 34% [76], and 33% [77] achieved by board-certified ophthalmologists performing indirect ophthalmoscopy and to reported sensitivities of 92% [78] and 89% [79] achieved by ophthalmologists interpreting digital fundus photographs. Our analysis suggests that the performance of ML algorithms in detecting DR based on color fundus photographs is likely to be on par with human clinicians and supports a previous study that compared humans head to head with ML. Rajalakshmi et al [37] compared the performance of an AI DR screening software (EyeArt) on smartphone-based fundus photographs of 296 patients to the performance of human graders who evaluated the same data set. The EyeArt achieved a high sensitivity of 95.8% for any retinopathy and 99.1% for VTDR, both of which were on par with human graders. Our pooled data suggest that ML techniques are more sensitive than specific in DR detection. It is unclear whether this is a reflection of the limitations of ML techniques for this clinical purpose or whether it is by design. It is possible that model developers of these studies chose optimal statistical thresholds that favored sensitivity over specificity. Regardless, the lower specificity should not pose a major issue, as false negatives are much more problematic than are false positives in the context of disease screening.

XSL•FO
RenderX

Furthermore, the major causes of false positives in retinal image interpretation, including inadequate image quality and artifacts [55], are modifiable with future improvement in image quality control.

To further facilitate direct clinical interpretations, we used Fagan nomograms to determine whether a patient with a positive or negative finding by ML actually has that particular finding as per the gold standard. For any DR, in a population with a DR prevalence of 36%, a positive likelihood ratio of 12.4 translates into a posttest probability of 87%. In other words, approximately 9 out of 10 patients with a positive ML diagnosis of DR can be expected to have DR as per the gold standard. The diagnostic value for ML to rule out DR performed as well as its rule-in value. In the same population, a negative ML diagnosis translates into a 4% posttest probability of any DR (negative likelihood ratio 0.07) and only a 1% posttest probability of mtmDR (negative likelihood ratio 0.05). These numbers again suggest that ML is extremely sensitive in detecting overall DR and mtmDR based on color fundus photographs and that the rate of false negatives are low.

We performed an in-depth analysis of studies that involved the detection of mtmDR, as Abramoff et al's [15] pivotal trial involved the detection of mtmDR and led to the FDA's approval of the first fully automated ML system. Among the 16 mtmDR studies conducted by other research teams that were also externally validated, 14 showed performance superior to the preset end points (sensitivity >85%; specificity > 82.5%) used in Abramoff et al's trial. Although only 5 out of these studies were prospectively evaluated in a real-world setting as the Abramoff algorithm was, this suggests that Abramoff et al's trial was no accident and that ML algorithms in general are likely capable of producing clinical grade detections of mtmDR based on color fundus photographs. In addition, no statistically significant difference in pooled AUROC between studies validated by benchmark databases and studies validated by clinical databases was identified within this group.

To the best of our knowledge, previous meta-reviews on DR screening have focused on the performance of DL algorithms alone [80,81]. DL is only a subtype of ML, and other ML techniques, such as SVM and RF, can be used to detect DR as well. Therefore, our meta-analysis was more comprehensive than these previous studies, as it included all ML studies, including DL studies, published through 2020. In addition, the review by Nielsen et al [80] did not conduct pooled analysis on the results of past studies, while our study did. The meta-analysis by Islam et al [81] focused mainly on detection of referable DR, while our study was broader and more fine grained, as it evaluated the ability of ML to detect different categories of DR, including any DR, mtmDR (referrable DR), VTDR, and PDR. These analyses are clinically meaningful, as different categories of DR require different management strategies. For example, while patients with moderate nonproliferative DR (a subset within referrable DR) should be further evaluated by ophthalmologists at some point, patients with VTDR require immediate referrals to retinal specialists.

Use and Predominance of NN Algorithms

NN algorithms, especially deep convolutional NN algorithms, were generally recognized as the best ML technique for automated medical image analysis. NN algorithms were also the most-used technique in diagnosing DR of all categories in our study, being used in 37 of the 60 (62%) studies. As for the 23 studies evaluating mtmDR (Table 2), NN algorithms were used in 22 studies and contributed to the high pooled AUROC of 0.98 (95% CI 0.96-0.99). In addition, we ranked the performance of the included studies by sensitivity, specificity, and quality. The top-5 performing, high-quality (based on QUADUS-2 and study design) studies are listed in Multimedia Appendix 10, and 4 out of the 5 studies used NN algorithms. This result confirms that NN is the cutting-edge ML technique for medical image classification, at least in the context of DR detection.

Limitations

Our study was based on a rigorous literature search, and a validated appraisal tool was used to determine the risk of bias of included studies. Several limitations should be considered, however. First, of the 60 studies included for final analysis, only 35 applied true external validation. For those studies without external validation, the generalizability of their ML algorithms was not adequately evaluated, and thus their reported performance should be interpreted with caution. Second, without sufficient details, we were unable to conduct subgroup analysis on populations with available key factors of DR that could influence the clinical practicability of the diagnostic tool. Bias could have been introduced by poor reporting of patient characteristics of the included studies. Finally, except for Abramoff et al's trial [15] and 5 other prospectively conducted studies [48,52,55,60,66], all other studies on ML-based DR diagnosis were validated by retrospective data. Due to spectrum bias, an overestimation of ML's performance in a real-world setting is possible and should be considered.

Conclusions

ML algorithms for diagnosing DR based on color fundus photographs have shown high diagnostic accuracy for different categories of DR. Specifically, the performances of ML algorithms in detecting mtmDR, the widely accepted threshold for clinically relevant DR, compare favorably to those of clinical examinations by ophthalmologists and to those of expert grading of digital fundus photographs. To the best of our knowledge, this is the first meta-analysis in the published literature that quantitatively assessed the performance of ML algorithms for a specific medical image classification task. As evidence-based medicine expands from therapy to diagnosis, the information from this systematic review may provide important evidence in the determination of the proper and efficacious use of ML algorithms in the diagnosis or screening of DR and may serve as a framework for similar analyses of other medical conditions conducted in the future. However, our meta-analysis also showed that a significant portion of the published studies had methodological flaws, such as the lack of external validation and presence of spectrum bias. Therefore, more rigorous prospective studies would be helpful in establishing the true efficacy of these algorithms in real-life clinical care.

XSL•FO
RenderX

Acknowledgments

This study was funded by the Taiwan National Ministry of Science and Technology (grants MOST 104-2314-B-002 -039 -MY3 and MOST 106-2811-B-002-048). The sponsor or funding organization had no role in the design or conduct of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search detail for PubMed and EMBASE.
[DOCX File , 33 KB - jmir_v23i7e23863_app1.docx]

Multimedia Appendix 2
Summary of characteristics of the included studies.
[DOCX File , 74 KB - jmir_v23i7e23863_app2.docx]

Multimedia Appendix 3
Characteristics of public databases for benchmarking diabetic retinopathy detection from color fundus photographs.
[DOCX File , 40 KB - jmir_v23i7e23863_app3.docx]

Multimedia Appendix 4
Principles of major machine learning algorithms and methods applied in automated image analysis.
[DOCX File , 39 KB - jmir_v23i7e23863_app4.docx]

Multimedia Appendix 5
Summary of QUADAS-2 assessment for included studies.
[PDF File (Adobe PDF File), 145 KB - jmir_v23i7e23863_app5.pdf]

Multimedia Appendix 6
Summary of data of included studies.
[DOCX File , 81 KB - jmir_v23i7e23863_app6.docx]

Multimedia Appendix 7
Fagan plot for diagnosis of different categories of diabetic retinopathy on color fundus photographs.
[PDF File (Adobe PDF File), 728 KB - jmir_v23i7e23863_app7.pdf]

Multimedia Appendix 8
Deeks funnel plot for 4 main types of diabetic retinopathy lesions on color fundus photograph.
[PDF File (Adobe PDF File), 374 KB - jmir_v23i7e23863_app8.pdf]

Multimedia Appendix 9
A filled funnel plot that filled the potential missing studies (square) due to publication bias.
[DOCX File , 264 KB - jmir_v23i7e23863_app9.docx]

Multimedia Appendix 10
Characteristics and best results of 5 high-quality studies with top performances.
[DOCX File , 46 KB - jmir_v23i7e23863_app10.docx]

References

1. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, American Diabetes Association. Diabetic retinopathy. Diabetes Care 2003 Jan;26(1):226-229. [doi: 10.2337/diacare.26.1.226] [Medline: 12502685]
2. Yau JWY, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 2012 Mar;35(3):556-564 [FREE Full text] [doi: 10.2337/dc11-1909] [Medline: 22301125]

XSL•FO
RenderX

3. Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. Indian J Ophthalmol 2012;60(5):428-431. [doi: 10.4103/0301-4738.100542] [Medline: 22944754]
4. Vashist P, Singh S, Gupta N, Saxena R. Role of early screening for diabetic retinopathy in patients with diabetes mellitus: an overview. Indian J Community Med 2011 Oct;36(4):247-252 [FREE Full text] [doi: 10.4103/0970-0218.91324] [Medline: 22279252]
5. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018 Feb;14(2):88-98. [doi: 10.1038/nrendo.2017.151] [Medline: 29219149]
6. Viswanath K, McGavin DDM. Diabetic retinopathy: clinical findings and management. Community Eye Health 2003;16(46):21-24 [FREE Full text] [Medline: 17491851]
7. Heitmar R, Kalitzeos AA, Patel SR, Prabhu-Das D, Cubbidge RP. Comparison of subjective and objective methods to determine the retinal arterio-venous ratio using fundus photography. J Optom 2015;8(4):252-257 [FREE Full text] [doi: 10.1016/j.optom.2014.07.002] [Medline: 26386537]
8. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Dec;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
9. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. IEEE Signal Process Mag 2010 Jul;27(4):25-38 [FREE Full text] [doi: 10.1109/MSP.2010.936730] [Medline: 25382956]
10. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. JAMA 2016 Dec 13;316(22):2366-2367 [FREE Full text] [doi: 10.1001/jama.2016.17563] [Medline: 27898977]
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
12. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
13. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
14. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
15. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1:39 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
16. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
17. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. Med Biol Eng Comput 2017 Nov;55(11):1959-1974. [doi: 10.1007/s11517-017-1638-6] [Medline: 28353133]
18. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016 Oct 01;57(13):5200-5206. [doi: 10.1167/iovs.16-19964] [Medline: 27701631]
19. Adal KM, Sidibé D, Ali S, Chaum E, Karnowski TP, Mériaudeau F. Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. Comput Methods Programs Biomed 2014 Apr;114(1):1-10. [doi: 10.1016/j.cmpb.2013.12.009] [Medline: 24529636]
20. Agurto C, Barriga ES, Murray V, Nemeth S, Crammer R, Bauman W, et al. Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. Invest Ophthalmol Vis Sci 2011 Jul 29;52(8):5862-5871. [doi: 10.1167/iovs.10-7075] [Medline: 21666234]
21. Annie GVG, Kaja MS. Diabetic retinopathy screening system: A validation analysis with multiple fundus image databases. Biomedical Research (India) 2017;28(4):A.
22. Chaum E, Karnowski TP, Govindasamy VP, Abdelrahman M, Tobin KW. Automated diagnosis of retinopathy by content-based image retrieval. Retina 2008;28(10):1463-1477. [doi: 10.1097/IAE.0b013e31818356dd] [Medline: 18997609]
23. Ganesan K, Martis RJ, Acharya UR, Chua CK, Min LC, Ng EYK, et al. Computer-aided diabetic retinopathy detection using trace transforms on digital fundus images. Med Biol Eng Comput 2014 Aug;52(8):663-672. [doi: 10.1007/s11517-014-1167-5] [Medline: 24958614]
24. Gardner GG, Keating D, Williamson TH, Elliott AT. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br J Ophthalmol 1996 Nov;80(11):940-944 [FREE Full text] [doi: 10.1136/bjo.80.11.940] [Medline: 8976718]
25. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology 2017 Dec;124(7):962-969. [doi: 10.1016/j.ophtha.2017.02.008] [Medline: 28359545]

XSL•FO
RenderX

26. Gupta G, Kulasekaran S, Ram K, Joshi N, Sivaprakasam M, Gandhi R. Local characterization of neovascularization and identification of proliferative diabetic retinopathy in retinal fundus images. Comput Med Imaging Graph 2017 Jan;55:124-132. [doi: 10.1016/j.compmedimag.2016.08.005] [Medline: 27634547]
27. Jelinek HF, Cree MJ, Leandro JJG, Soares JVB, Cesar RM, Luckie A. Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy. J Opt Soc Am A Opt Image Sci Vis 2007 May;24(5):1448-1456. [doi: 10.1364/josaa.24.001448] [Medline: 17429492]
28. Malathi K, Nedunchelian R. A recursive support vector machine (RSVM) algorithm to detect and classify diabetic retinopathy in fundus retina images. biomedicalresearch 2018:57-64. [doi: 10.4066/biomedicalresearch.29-16-2328]
29. Fadafen MK, Mehrshad N, Razavi SM. Detection of diabetic retinopathy using computational model of human visual system. biomedicalresearch 2018;29(9):1956-1960. [doi: 10.4066/biomedicalresearch.29-18-551]
30. Li Z, Keel S, Liu C, He Y, Meng W, Scheetz J, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. Diabetes Care 2018 Dec;41(12):2509-2516. [doi: 10.2337/dc18-0147] [Medline: 30275284]
31. Bala MP, Vijayachitra S. Extraction of retinal blood vessels and diagnosis of proliferative diabetic retinopathy using extreme learning machine. J Med Imaging Hlth Inform 2015 Apr 01;5(2):248-256. [doi: 10.1166/jmihi.2015.1380]
32. Orlando JI, Prokofyeva E, Del Fresno M, Blaschko MB. An ensemble deep learning based approach for red lesion detection in fundus images. Comput Methods Programs Biomed 2018 Jan;153:115-127. [doi: 10.1016/j.cmpb.2017.10.017] [Medline: 29157445]
33. Orlando JI, van Keer K, Barbosa Breda J, Manterola HL, Blaschko MB, Clausse A. Proliferative diabetic retinopathy characterization based on fractal features: Evaluation on a publicly available dataset. Med Phys 2017 Dec;44(12):6425-6434. [doi: 10.1002/mp.12627] [Medline: 29044550]
34. Pires R, Carvalho T, Spurling G, Goldenstein S, Wainer J, Luckie A, et al. Automated multi-lesion detection for referable diabetic retinopathy in indigenous health care. PLoS One 2015;10(6):e0127664 [FREE Full text] [doi: 10.1371/journal.pone.0127664] [Medline: 26035836]
35. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. Med Image Anal 2017 Jul;39:178-193 [FREE Full text] [doi: 10.1016/j.media.2017.04.012] [Medline: 28511066]
36. Quellec G, Lamard M, Abràmoff MD, Decencière E, Lay B, Erginay A, et al. A multiple-instance learning framework for diabetic retinopathy screening. Med Image Anal 2012 Aug;16(6):1228-1240. [doi: 10.1016/j.media.2012.06.003] [Medline: 22850462]
37. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) 2018 Jun;32(6):1138-1144 [FREE Full text] [doi: 10.1038/s41433-018-0064-9] [Medline: 29520050]
38. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a deep learning algorithm for automatic diagnosis of diabetic retinopathy. Stud Health Technol Inform 2017;245:559-563. [Medline: 29295157]
39. Ramachandran N, Hong SC, Sime MJ, Wilson GA. Diabetic retinopathy screening using deep neural network. Clin Exp Ophthalmol 2018 May;46(4):412-416. [doi: 10.1111/ceo.13056] [Medline: 28881490]
40. Sangeethaa SN, Uma Maheswari P. An intelligent model for blood vessel segmentation in diagnosing DR using CNN. J Med Syst 2018 Aug 15;42(10):175. [doi: 10.1007/s10916-018-1030-6] [Medline: 30109508]
41. B. Sumathy, S. Poornachandra. Automated DR and prediction of various related diseases of retinal fundus images. biomedicalresearch 2018:325-332. [doi: 10.4066/biomedicalresearch.29-17-480]
42. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
43. Usman Akram M, Khalid S, Tariq A, Younus Javed M. Detection of neovascularization in retinal images using multivariate m-Mediods based classifier. Comput Med Imaging Graph 2013;37(5-6):346-357 [FREE Full text] [doi: 10.1016/j.compmedimag.2013.06.008] [Medline: 23916066]
44. Welikala RA, Fraz MM, Dehmeshki J, Hoppe A, Tah V, Mann S, et al. Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Comput Med Imaging Graph 2015 Jul;43:64-77. [doi: 10.1016/j.compmedimag.2015.03.003] [Medline: 25841182]
45. Yu S, Xiao D, Kanagasingam Y. Machine learning based automatic neovascularization detection on optic disc region. IEEE J. Biomed. Health Inform 2018 May;22(3):886-894. [doi: 10.1109/jbhi.2017.2710201]
46. Zhang Y, An M. An active learning classifier for further reducing diabetic retinopathy screening system cost. Comput Math Methods Med 2016;2016:4345936 [FREE Full text] [doi: 10.1155/2016/4345936] [Medline: 27660645]
47. Badgujar R, Deore P. Hybrid nature inspired SMO-GBM classifier for exudate classification on fundus retinal images. IRBM 2019 Mar;40(2):69-77 [FREE Full text] [doi: 10.1016/j.irbm.2019.02.003] [Medline: 32595044]
48. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MYT, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. The Lancet Digital Health 2019 May;1(1):e35-e44. [doi: 10.1016/s2589-7500(19)30004-4]

XSL•FO
RenderX

49. Bhaskaranand M, Ramachandra C, Bhat S, Cuadros J, Nittala MG, Sadda SR, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. Diabetes Technol Ther 2019 Nov;21(11):635-643 [FREE Full text] [doi: 10.1089/dia.2019.0164] [Medline: 31335200]
50. Chowdhury AR, Chatterjee T, Banerjee S. A Random Forest classifier-based approach in the detection of abnormalities in the retina. Med Biol Eng Comput 2019 Jan;57(1):193-203. [doi: 10.1007/s11517-018-1878-0] [Medline: 30076537]
51. Colomer A, Igual J, Naranjo V. Detection of early signs of diabetic retinopathy based on textural and morphological information in fundus images. Sensors (Basel) 2020 Feb 13;20(4):1005 [FREE Full text] [doi: 10.3390/s20041005] [Medline: 32069912]
52. Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol 2019 Jun 13:987-993 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.2004] [Medline: 31194246]
53. He J, Cao T, Xu F, Wang S, Tao H, Wu T, et al. Artificial intelligence-based screening for diabetic retinopathy at community hospital. Eye (Lond) 2020 Mar;34(3):572-576 [FREE Full text] [doi: 10.1038/s41433-019-0562-4] [Medline: 31455902]
54. Hemanth DJ, Anitha J, Son LH, Mittal M. Diabetic retinopathy diagnosis from retinal images using modified Hopfield neural network. J Med Syst 2018 Oct 31;42(12):247. [doi: 10.1007/s10916-018-1111-6] [Medline: 30382410]
55. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018 Sep 07;1(5):e182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]
56. Khojasteh P, Passos Júnior LA, Carvalho T, Rezende E, Aliahmad B, Papa JP, et al. Exudate detection in fundus images using deeply-learnable features. Comput Biol Med 2019 Jan;104:62-69 [FREE Full text] [doi: 10.1016/j.compbiomed.2018.10.031] [Medline: 30439600]
57. Li F, Liu Z, Chen H, Jiang M, Zhang X, Wu Z. Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. Transl Vis Sci Technol 2019 Nov;8(6):4 [FREE Full text] [doi: 10.1167/tvst.8.6.4] [Medline: 31737428]
58. Long S, Huang X, Chen Z, Pardhan S, Zheng D. automatic detection of hard exudates in color retinal images using dynamic threshold and SVM classification: algorithm development and evaluation. Biomed Res Int 2019;2019:3926930 [FREE Full text] [doi: 10.1155/2019/3926930] [Medline: 30809539]
59. Natarajan S, Jain A, Krishnan R, Rogye A, Sivaprasad S. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. JAMA Ophthalmol 2019 Aug 08:1182-1188 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.2923] [Medline: 31393538]
60. Nazir T, Irtaza A, Shabbir Z, Javed A, Akram U, Mahmood MT. Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines. Artif Intell Med 2019 Aug;99:101695 [FREE Full text] [doi: 10.1016/j.artmed.2019.07.003] [Medline: 31606114]
61. Pires R, Avila S, Wainer J, Valle E, Abramoff MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. Artif Intell Med 2019 May;96:93-106 [FREE Full text] [doi: 10.1016/j.artmed.2019.03.009] [Medline: 31164214]
62. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med 2019;2:25 [FREE Full text] [doi: 10.1038/s41746-019-0099-8] [Medline: 31304372]
63. Riaz H, Park J, Choi H, Kim H, Kim J. Deep and densely connected networks for classification of diabetic retinopathy. Diagnostics 2020 Jan 02;10(1):24. [doi: 10.3390/diagnostics10010024]
64. Shah P, Mishra DK, Shanmugam MP, Doshi B, Jayaraj H, Ramanjulu R. Validation of Deep Convolutional Neural Network-based algorithm for detection of diabetic retinopathy - Artificial intelligence versus clinician for screening. Indian J Ophthalmol 2020 Feb;68(2):398-405 [FREE Full text] [doi: 10.4103/ijo.IJO_966_19] [Medline: 31957737]
65. Son J, Shin JY, Kim HD, Jung K, Park KH, Park SJ. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. Ophthalmology 2020 Jan;127(1):85-94 [FREE Full text] [doi: 10.1016/j.ophtha.2019.05.029] [Medline: 31281057]
66. Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, et al. Simple, Mobile-based Artificial Intelligence Algorithm in the detection of Diabetic Retinopathy (SMART) study. BMJ Open Diabetes Res Care 2020 Jan;8(1):e000892 [FREE Full text] [doi: 10.1136/bmjdrc-2019-000892] [Medline: 32049632]
67. Stevenson CH, Hong SC, Ogbuehi KC. Development of an artificial intelligence system to classify pathology and clinical features on retinal fundus images. Clin Exp Ophthalmol 2019 May;47(4):484-489. [doi: 10.1111/ceo.13433] [Medline: 30370587]
68. Ullah H, Saba T, Islam N, Abbas N, Rehman A, Mehmood Z, et al. An ensemble classification of exudates in color fundus images using an evolutionary algorithm based optimal features selection. Microsc Res Tech 2019 Jan 24;82(4):361-372. [doi: 10.1002/jemt.23178]
69. Verbraak FD, Abramoff MD, Bausch GCF, Klaver C, Nijpels G, Schlingemann RO, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. Diabetes Care 2019 Apr;42(4):651-656. [doi: 10.2337/dc18-0148] [Medline: 30765436]

XSL•FO
RenderX

70. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. PLoS One 2019;14(6):e0217541 [FREE Full text] [doi: 10.1371/journal.pone.0217541] [Medline: 31170223]
71. Wang H, Yuan G, Zhao X, Peng L, Wang Z, He Y, et al. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening. Comput Methods Programs Biomed 2020 Jul;191:105398 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105398] [Medline: 32092614]
72. Xie L, Yang S, Squirrell D, Vaghefi E. Towards implementation of AI in New Zealand national diabetic screening program: Cloud-based, robust, and bespoke. PLoS ONE 2020 Apr 10;15(4):e0225015. [doi: 10.1371/journal.pone.0225015]
73. Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. Comput Biol Med 2020 Jan;116:103537 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103537] [Medline: 31747632]
74. Yang W, Zheng B, Wu M, Zhu S, Fei F, Weng M, et al. An evaluation system of fundus photograph-based intelligent diagnostic technology for diabetic retinopathy and applicability for research. Diabetes Ther 2019 Oct;10(5):1811-1822 [FREE Full text] [doi: 10.1007/s13300-019-0652-0] [Medline: 31290125]
75. Lawrence MG. The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: an analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. Trans Am Ophthalmol Soc 2004;102:321-340 [FREE Full text] [Medline: 15747766]
76. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. Am J Ophthalmol 2002 Aug;134(2):204-213. [doi: 10.1016/s0002-9394(02)01522-2] [Medline: 12140027]
77. Pugh JA, Jacobson JM, Van Heuven WA, Watters JA, Tuley MR, Lairson DR, et al. Screening for diabetic retinopathy. The wide-angle retinal camera. Diabetes Care 1993 Jun;16(6):889-895. [doi: 10.2337/diacare.16.6.889] [Medline: 8100761]
78. Lopez-Bastida J, Cabrera-Lopez F, Serrano-Aguilar P. Sensitivity and specificity of digital retinal imaging for screening diabetic retinopathy. Diabet Med 2007 Apr;24(4):403-407. [doi: 10.1111/j.1464-5491.2007.02074.x] [Medline: 17298591]
79. Saari JM, Summanen P, Kivelä T, Saari KM. Sensitivity and specificity of digital retinal images in grading diabetic retinopathy. Acta Ophthalmol Scand 2004 Apr;82(2):126-130 [FREE Full text] [doi: 10.1111/j.1600-0420.2004.00240.x] [Medline: 15043527]
80. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol Retina 2019 Apr;3(4):294-304. [doi: 10.1016/j.oret.2018.10.014] [Medline: 31014679]
81. Islam MM, Yang H, Poly TN, Jian W, Jack Li Y. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. Comput Methods Programs Biomed 2020 Jul;191:105320 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105320] [Medline: 32088490]

Abbreviations

AI: artificial intelligence
AUROC: area under the receiver operating characteristics
DL: deep learning
DR: diabetic retinopathy
DRIVE: Digital Retinal Images for Vessel Extraction
FDA: US Food and Drug Administration
ML: machine learning
mtmDR: more-than-mild diabetic retinopathy
NN: neural network
OR: odds ratio
PDR: proliferative diabetic retinopathy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2
RF: random forest
ROC: receiver operating characteristics
STARE: Structured Analysis of the Retina
SVM: support vector machine
VTDR: vision-threatening diabetic retinopathy

XSL•FO
RenderX

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by G Lim, K Koshechkin; comments to author 06.11.20; revised version received 19.11.20; accepted 30.04.21; published 05.07.21.

Please cite as:
Wu JH, Liu TYA, Hsu WT, Ho JHC, Lee CC
Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis
J Med Internet Res 2021;23(7):e23863
URL: https://www.jmir.org/2021/7/e23863
doi:10.2196/23863
PMID:34407500

©Jo-Hsuan Wu, T Y Alvin Liu, Wan-Ting Hsu, Jennifer Hui-Chun Ho, Chien-Chang Lee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-analysis

Scarlet Nazarian[1*], BSc, MBBS, MRCS; Ben Glover[1*], MBBS, MRCP; Hutan Ashrafian[1*], PhD; Ara Darzi[1*], PhD; Julian Teare[1*], PhD

Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Hutan Ashrafian, PhD
Department of Surgery and Cancer
Imperial College London
10th Floor, QEQM Building, St Mary's Hospital
Praed Street
London, W2 1NY
United Kingdom
Phone: 44 2075895111
Email: h.ashrafian@imperial.ac.uk

Abstract

Background: Colonoscopy reduces the incidence of colorectal cancer (CRC) by allowing detection and resection of neoplastic polyps. Evidence shows that many small polyps are missed on a single colonoscopy. There has been a successful adoption of artificial intelligence (AI) technologies to tackle the issues around missed polyps and as tools to increase the adenoma detection rate (ADR).

Objective: The aim of this review was to examine the diagnostic accuracy of AI-based technologies in assessing colorectal polyps.

Methods: A comprehensive literature search was undertaken using the databases of Embase, MEDLINE, and the Cochrane Library. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed. Studies reporting the use of computer-aided diagnosis for polyp detection or characterization during colonoscopy were included. Independent proportions and their differences were calculated and pooled through DerSimonian and Laird random-effects modeling.

Results: A total of 48 studies were included. The meta-analysis showed a significant increase in pooled polyp detection rate in patients with the use of AI for polyp detection during colonoscopy compared with patients who had standard colonoscopy (odds ratio [OR] 1.75, 95% CI 1.56-1.96; $P<.001$). When comparing patients undergoing colonoscopy with the use of AI to those without, there was also a significant increase in ADR (OR 1.53, 95% CI 1.32-1.77; $P<.001$).

Conclusions: With the aid of machine learning, there is potential to improve ADR and, consequently, reduce the incidence of CRC. The current generation of AI-based systems demonstrate impressive accuracy for the detection and characterization of colorectal polyps. However, this is an evolving field and before its adoption into a clinical setting, AI systems must prove worthy to patients and clinicians.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020169786; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020169786

(J Med Internet Res 2021;23(7):e27370) doi:10.2196/27370

KEYWORDS

artificial intelligence; colonoscopy; computer-aided diagnosis; machine learning; polyp

XSL•FO
RenderX

Introduction

Colorectal cancer (CRC) is the third-leading malignancy worldwide and a leading cause of mortality [1]. CRC typically develops from sporadic colorectal adenomatous polyps, and colonoscopy is established for the detection and resection of these lesions, which has been shown to reduce the incidence and mortality from CRC [2]. However, as with any procedure, endoscopic polyp detection has operator-dependent limitations. There is evidence highlighting that small polyps may be missed at colonoscopy with a miss rate for adenomas as high as 26% [3]. The primary colonoscopy quality indicator is the adenoma detection rate (ADR). Given that ADR is inversely proportional to postcolonoscopy CRC risk, with each 1% increase in ADR equivalent to a 3% decrease in the subsequent risk of cancer [4], there is an unmet need to tackle the problems that prevent high-quality colonoscopy.

Human and technical factors lead to a small but significant proportion of missed polyps during colonoscopy. Several studies have suggested that ADR can be increased by improving the educational and behavioral skills of the endoscopist. Training programs, consisting of hands-on teaching and regular feedback, showed good results in increasing ADR in trials [5,6]. However, the increase in detection from baseline in these studies was minimal and the ability of even expert endoscopists to detect very small, subtle, or flat lesions remains a limiting factor.

Recently, there has been a successful adoption of artificial intelligence (AI) technologies in health care diagnostics [7]. The ability of AI, specifically machine learning approaches, to differentiate and characterize distinct pathologies is continuously enhancing early computer-aided diagnosis (CAD) techniques. Deep learning models are built using artificial neural networks and have proven very useful with analysis of big data in health care. Convolutional neural networks (CNNs) and their variants with AI models have become the most preferred and widely used methods in medical image analysis. Convolutional layers convolve the input and pass its result to the next layer. Application of AI in colonoscopy has focused more on polyp detection than characterization, driven by the development of deep CNNs (DCNNs). The architecture of these algorithms includes multiple layers of processing between the input and output layers, allowing analysis of complex data with efficient performance. The most advanced polyp detection systems are those that can be applied to video-based analysis during colonoscopy.

In the field of endoscopy, a machine learning algorithm can be trained to recognize or characterize polyps in real time. Two endoscopic approaches have been studied: techniques used for analysis of nonmagnified endoscopic images and those for cellular imaging at a microscopic level (ie, optical biopsy).

The idea of such approaches is that by detecting more polyps (ie, increasing the polyp detection rate [PDR]), there will be a corresponding reduction in the number of missed adenomas and, consequently, a reduction in the subsequent risk of CRC. However, this presents a financial burden on health care systems, especially the histopathology departments, involved in analysis of resected tissue, which will only increase with the increase in detection of polyps. The ultimate goal of a CAD system would be the reliable detection of every polyp within the colon during the colonoscopy procedure, while also characterizing them as hyperplastic or adenomatous to guide decision making for polypectomy and histopathological examination [8]. The Preservation and Incorporation of Valuable endoscopic Innovations (PIVI) initiative, set by the American Society of Gastrointestinal Endoscopy (ASGE), has established a desired threshold for the introduction of new endoscopic technologies, including the optical diagnosis of diminutive colorectal polyps [9]. Despite several, predominantly single-site, studies meeting the PIVI criteria showing that a “resect and discard” strategy or a “diagnose and leave” strategy could be adopted [10,11], a recent multicenter study showed that the accuracy of optical diagnosis requires imaging advances before it can be used to determine surveillance without histology [12].

Machine learning by definition is a model that is able to constantly adapt and improve when presented with new information. To ensure this refinement, large quantities of good-quality data should be used for training the algorithm. Current AI systems that are not synthesized in this way are prone to the risk of *overfitting*, whereby the system performs well with training data to the extent that it negatively impacts its performance when tested on new data [13]. Thus, for an AI system to be successful in its ability to detect and characterize polyps, it should adopt a machine learning model based on good-quality high-yield data and the model should have a high sensitivity for the detection of polyps, have a low rate of false positives, and be able to maintain fast processing speeds to be applicable in near-real time during colonoscopy [14].

Our aims were to systematically review and meta-analyze the diagnostic accuracy of AI-based technologies in the detection and characterization of colorectal polyps.

Methods

This review was carried out and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [15]. It has been registered on PROSPERO (International Prospective Register of Systematic Reviews) (registration No. CRD42020169786).

Search Strategy

A comprehensive literature search was undertaken using the databases of Embase, MEDLINE, and the Cochrane Library. All published articles up until October 2020 were included. Search terms used in Embase and MEDLINE included “colon*,” “polyp,” “artificial intelligence OR machine learning,” and “computer aided or assisted and diagnos* OR detect*.” Studies in the Cochrane Library were identified with the terms “colonic polyp,” “artificial intelligence,” and “diagnosis, computer-assisted” (Multimedia Appendix 1).

Inclusion and Exclusion Criteria

Inclusion criteria were as follows:

- Studies reporting computer-aided detection of colorectal polyps retrospectively, using endoscopic images or videos

XSL•FO
RenderX

- Studies reporting computer-aided classification of colorectal polyps retrospectively, using endoscopic images or videos
- Studies reporting the use of CAD of colorectal polyps during colonoscopy
- Studies reporting ADR, PDR, sensitivity, specificity, and diagnostic accuracy data or studies with adequate information to calculate these data
- Studies published or translated into English.

Exclusion criteria were as follows:

- Studies with no original data present (eg, review article or letter)
- Studies with no full text available
- Studies conducted in patients with inflammatory bowel disease (IBD)
- Studies greater than 20 years old
- Studies without adequate data to calculate sensitivity, specificity, and diagnostic accuracy data; PDR and ADR; adenoma miss rate; or mean adenomas per patient, or those not reporting these data.

Study Selection

The retrieved articles were screened for duplicates by two reviewers; these were excluded. Titles and abstracts were then screened for relevance by two reviewers independently, and irrelevant studies were excluded. Following this, full-text reviews of remaining studies were completed. The reference lists of identified review articles and included papers were scrutinized for relevant studies. Disagreements about eligibility were settled by consensus, both after screening and following full‐text review. Inclusion and exclusion criteria were met by all final articles.

Data Extraction

Data were gathered from studies and placed onto a standard spreadsheet template. For each study, we extracted the following data: study details (ie, first author, year of publication, and journal), primary outcome (ie, polyp detection vs characterization), study design (ie, type of study, method of AI, and exclusion criteria), information on type of imaging modality (ie, images or videos, images for training, and images for validation), and information regarding diagnostic accuracy characteristics (ie, sensitivity, specificity, accuracy, ADR, and PDR).

Study Quality Assessment

Study quality was independently assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [16]. Each domain was classified as low-risk, high-risk, or unclear risk of bias. For randomized controlled trials (RCTs), the Jadad scale was used for quality scoring [17]. Studies with a Jadad score of 3 or more were considered *good* quality.

Statistical Analysis

Independent proportions and their differences were calculated and pooled through DerSimonian and Laird random-effects modeling. This considered both between-study and within-study variances, which contributed to study weighting. Pooled values and 95% CIs were computed and represented on forest plots. Statistical heterogeneity was determined by the I^2 statistic, where <30% was low, 30%-60% was moderate, and >60% was high. Analyses were performed using Stata, version 15 (StataCorp LLC). Probability values of $P \leq .05$ were considered statistically significant.

Results

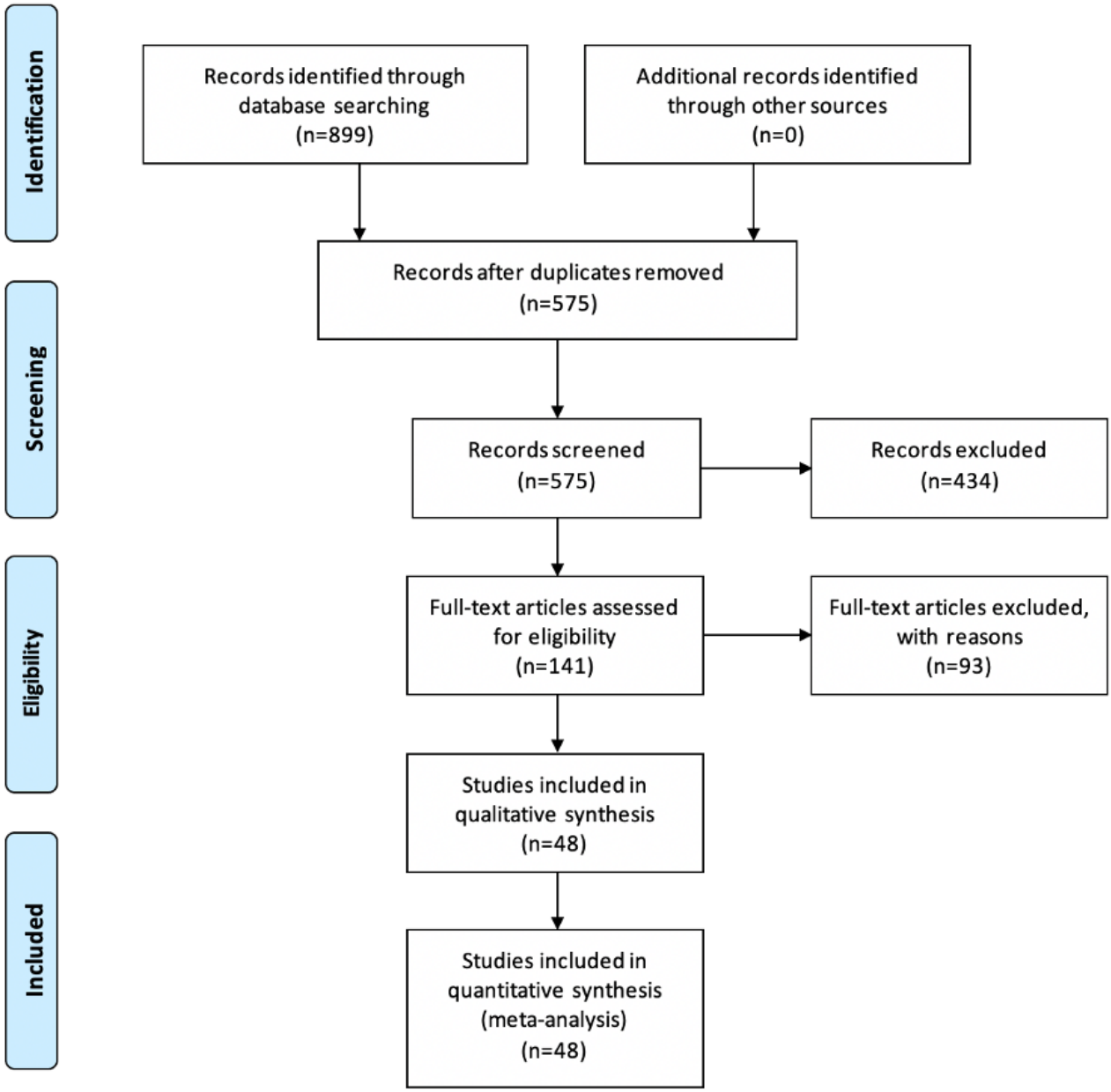
Search Results and Characteristics

A total of 899 articles were identified from the database searches. After removing duplicates, 575 records were screened on the basis of titles and abstracts. A total of 141 articles were identified as appropriate for full-text review. Further evaluation and application of the exclusion criteria revealed 48 studies, which were included in this systematic review and meta-analysis. The study screening and selection process is shown in Figure 1.

Studies in this systematic review included preclinical studies for polyp detection (Table 1 [18-35]), preclinical studies for polyp characterization (Table 2 [11,13,36-55]), and recent RCTs (Table 3 [56-63]). The studies were all published between 2003 and 2020. The outcome measures were polyp detection in 18 studies, polyp characterization in 22 studies, and PDR in 8 studies. The studies analyzing sensitivity, specificity, and accuracy when testing each AI system were found to present results at the per-patient, per-polyp, and/or per-image levels, whereas the RCTs evaluating the ADR and PDR consistently presented per-patient results.

XSL•FO
RenderX

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



Studies for polyp detection predominantly used CNN or DCNN as their machine learning approach. A total of 14 studies for polyp detection were carried out retrospectively. There was a large variation in the number of images used by each paper to train or validate the AI systems in detecting polyps, with one study using 8 images [20] to train the system, while another used 5545 images [25].

In the majority of studies, narrow band imaging (NBI) or endocytoscopy was the imaging method of choice for characterizing polyps, with one exception in which the imaging modality was not stated [47]. Data for polyp characterization was gathered retrospectively in 18 studies. In 3 studies that collected data prospectively, a support vector machine classifier was used as the machine learning approach. Similarly to studies for polyp detection, those analyzing polyp characterization had a large variation in number of images used for training or validating the AI system. However, studies for polyp characterization focused more on the number of polyps used than on overall images, as seen in Table 2.

XSL•FO
RenderX

Table 1. Characteristics of included studies whose primary outcome was polyp detection.

Authors	Year	Recruitment	Machine learning approach	Imaging modality	Patients, n	Polyps, n	Total images, n	Images for training, n	Images for validation, n
Karkanis et al [18]	2003	Retrospective	CWC[a]	RGB[b]–color frame grabber	66	95	1380	180	1200
Fu et al [19]	2014	Retrospective	SFFS[c] with SVM[d] classifier	Still image enhanced by PCT[e]	100	—[f]	365	292	73
Wang et al [20]	2015	Retrospective	Polyp edge detection—ECSP[g]	Video clip	—	43	—	8	53
Tajbakhsh et al [21]	2015	Retrospective	Hybrid context-shape approach	CVC[h]-Colon database	—	15	300	300	—
Tajbakhsh et al [21]	2015	Retrospective	Hybrid context-shape approach	ASU[i]-Mayo database	—	10	19,400	300	—
Fernández-Esparrach et al [22]	2016	Retrospective	WM-DOVA[j] maps	White light colonoscope	—	31	612	—	—
Park and Sargent [23]	2016	Retrospective	CNN[k]-CRF[l] model	White light and NBI[m]	—	92	11,802	—	—
Urban et al [24]	2018	Retrospective	DCNN[n]	NBI images	>2000	—	8641	—	—
Wang et al [25]	2018	Retrospective	ANN[o]- SegNet architecture	Still images	2428	—	—	5545	27,113
Misawa et al [26]	2018	Retrospective	CNN	White light images	73	155	546	411	135
Figueiredo et al [27]	2019	Retrospective	SVM binary classifier	White light images	42	42	—	—	—
Yamada et al [28]	2019	Retrospective	Faster R-CNN[p]	Still images	—	752	—	>4000	4840
Becq et al [29]	2020	Prospective	ANN- SegNet architecture	Video	50	165	—	—	—
Gao et al [30]	2020	Retrospective	CNN	White light images	—	—	1709	1196	256
Guo et al [31]	2020	Retrospective	CNN-YOLO[q]	Video	283	—	1991	—	—
Lee et al [32]	2020	Prospective	CNN-YOLO	Video	15	26	—	8495	110,728
Ozawa et al [33]	2020	Retrospective	CNN	NBI and white light	12,895	309	—	20,431	7077
Misawa et al [34]	2020	Prospective	CNN-YOLO	White light images	1405	100	56,668	51,889	4769
Poon et al [35]	2020	Prospective	CNN-ResNet50, YOLO	Video	144	128	—	198,138	34,469

[a]CWC: color wavelet covariance.

[b]RGB: red, green, and blue.

[c]SFFS: sequential floating-forward selection.

[d]SVM: support vector machine.

[e]PCT: principal components transformation.

[f]This value was not reported.

[g]ECSP: edge cross-section profiles.

[h]CVC: Computer Vision Center.

[i]ASU: Arizona State University.

[j]WM-DOVA: window median depth of valleys accumulation.

[k]CNN: convolutional neural network.

XSL•FO
RenderX

[l]CRF: conditional random field.

[m]NBI: narrow band imaging.

[n]DCNN: deep convolutional neural network.

[o]ANN: artificial neural network.

[p]R-CNN: region-based convolutional neural network.

[q]YOLO: you only look once.

XSL•FO
RenderX

Table 2. Characteristics of included studies whose primary outcome was polyp characterization.

Authors	Year	Recruitment	Machine learning approach	Image modality	Patients, n	Polypsor lesions, n	Total images, n	Images for training, n	Images for validation, n
Tischendorf et al [36]	2010	Prospective pilot	SVM[a] classifier	Magnification NBI[b]	223	209	—[c]	208	—
Gross et al [37]	2011	Prospective	SVM classifier	Magnification NBI	214	434	—	433	—
Ganz et al [13]	2012	Retrospective	Shape-UCM[d]	NBI	—	—	—	58	87
Takemura et al [38]	2012	Retrospective	SVM classifier	Magnification NBI	—	371	—	1519	371
Mori et al [39]	2015	Retrospective	EC[e]-CAD[f]	EC	152	176	—	—	—
Kominami et al [11]	2016	Retrospective	SVM classifier	Magnification NBI	41	118	—	2247	—
Misawa et al [40]	2016	Retrospective	EndoBRAIN[g]	NBI and EC	—	85	1079	979	100
Mesejo et al [41]	2016	Retrospective	SfM[h]	White light and NBI	—	76	—	—	—
Mori et al [42]	2016	Retrospective	SVM classifier	EC-CAD	123	205	—	—	6051
Takeda et al [43]	2017	Retrospective	SVM classifier	EC-CAD	242	375	5843	5643	200
Byrne et al [44]	2017	Retrospective	DCNN[i]	NBI	—	125	—	60,089	—
Komeda et al [45]	2017	Retrospective	CNN	Endoscopic images	—	—	1200	—	—
Misawa et al [46]	2017	Retrospective	EndoBRAIN and ECV[j]-CAD	NBI	100	124	1834	173	1661
Mori et al [47]	2018	Retrospective	—	EC	—	144	—	—	—
Chen et al [48]	2018	Prospective	DNN[k]	NBI	193	284	2441	2157	284
Renner et al [49]	2018	Retrospective	DNN	NBI and HD-WL[l]	250	231	788	602	186
Mori et al [50]	2018	Prospective	SVM classifier	NBI and EC	325	466	—	61,925	450
Mori et al [50]	2018	Prospective	SVM classifier	NBI and EC	325	466	—	61,925	450
Kudo et al [51]	2019	Retrospective	EndoBRAIN system	White light, NBI, and EC	89	100	—	69,142	5065
Kudo et al [51]	2019	Retrospective	EndoBRAIN system	White light, NBI, and EC	89	100	—	69,142	5065
Figueiredo et al [52]	2019	Retrospective	Segmentation algorithm	NBI	10	11	86	43	43
Rodriguez-Diaz et al [53]	2020	Retrospective	DeepLab framework	High magnification NBI	286	607	740	—	—
Yang et al [54]	2020	Retrospective	CNN-Inception-ResNet	White light	1339	—	3828	—	240
Zachariah et al [55]	2020	Retrospective	CNN-Inception-ResNet	NBI and white light	—	—	6223	—	634

[a]SVM: support vector machine.

[b]NBI: narrow band imaging.

[c]This value was not reported.

[d]Shape-UCM is an algorithm for automatic polyp segmentation.

[e]EC: endocytoscopy.

[f]CAD: computer-aided diagnosis.

[g]EndoBRAIN is a novel artificial intelligence system.

XSL•FO
RenderX

[h]SfM: structure from motion.

[i]DCNN: deep convolutional neural network.

[j]ECV: endocytoscopic vascular pattern.

[k]DNN: deep neural network.

[l]HDWL: high-definition white light.

Table 3. Characteristics of randomized controlled trials whose primary outcome was polyp detection.

Authors	Year	Recruitment	Machine learning approach	Imaging modality	Patients, n	Polyps, n	PDR[a] –AI[b], %	PDR –control, %	ADR[c] –AI, %	ADR –control, %	Withdrawal time[d]; AI vs control, min	*P* value
Wang et al [56]	2019	Real-time, prospective	ANN[e]-SegNet architecture	Video stream	1058	767	45.02	29.10	29.12	20.34	6.18 vs 6.07	.15
Wang et al [57]	2020	Prospective	ANN-SegNet architecture	Video stream	962	809	52	37	34	28	6.48 vs 6.37	.14
Su et al [58]	2020	Prospective	DCNN[f]	Video stream	623	273	38.31	25.40	28.90	16.50	7.03 vs 5.68	<.001
Gong et al [59]	2020	Prospective	DCNN	Video stream	704	—[g]	47	34	16	8	6.38 vs 4.76	<.001
Liu et al [60]	2020	Prospective	ANN	Video stream	1026	734	43.65	27.81	39.10	23.89	6.82 vs 6.74	<.001
Luo et al [61]	2020	Prospective	CNN-YOLO[h]	Video stream	150	185	38.7	34.0	—	—	6.22 vs 6.17	.10
Repici et al [62]	2020	Prospective	CNN-GI Genius[i]	Video stream	685	596	—	—	54.8	40.4	6.95 vs 7.25	.10
Wang et al [63]	2020	Prospective	ANN-Endoscreener	Video stream	369	—	63.59	55.14	42.39	35.68	6.55 vs 6.51	.75

[a]PDR: polyp detection rate.

[b]AI: artificial intelligence.

[c]ADR: adenoma detection rate.

[d]Withdrawal time excluded the time to perform the biopsy.

[e]ANN: artificial neural network.

[f]DCNN: deep convolutional neural network.

[g]This value was not reported.

[h]YOLO: you only look once.

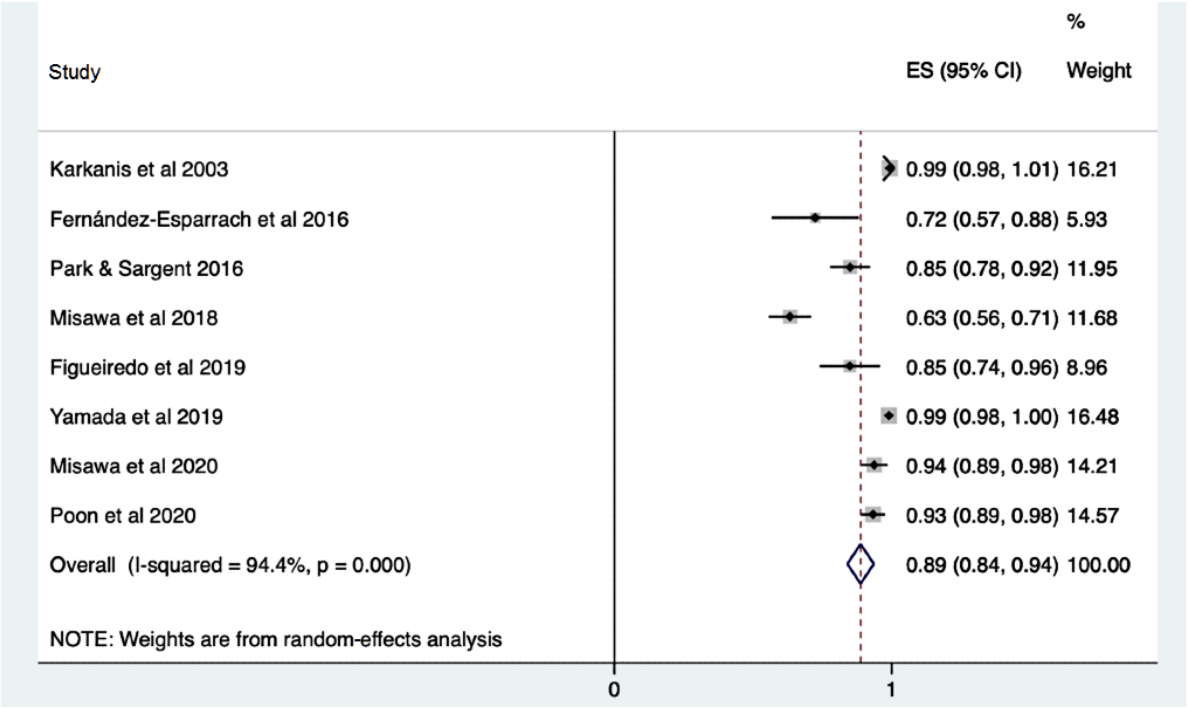
[i]GI Genius (Medtronic) is novel artificial intelligence system.

Detection or Localization of a Polyp

The diagnostic accuracy of the machine learning systems for detecting polyps was assessed using 103,049 still images in 10 studies, reporting a pooled sensitivity of 0.84 (95% CI 0.74-0.93), a specificity of 0.87 (95% CI 0.83-0.90), and an accuracy of 0.89 (95% CI 0.81-0.97). Lesions within video frames or images were used by 14 studies to report the diagnostic performance of their detection systems, highlighting a sensitivity of 0.92 (95% CI 0.89-0.95), a specificity of 0.89 (95% CI 0.84-0.94; Figure 2), and an accuracy of 0.87 (95% CI 0.76-0.97). There were 11 studies analyzing the accuracy of polyp detection through the use of images or video clips gathered from more than 17,401 patients. These demonstrated a sensitivity of 0.92 (95% CI 0.90-0.94), a specificity of 0.93 (95% CI 0.91-0.96), and accuracy of 0.92 (95% CI 0.87-0.98).

XSL•FO
RenderX

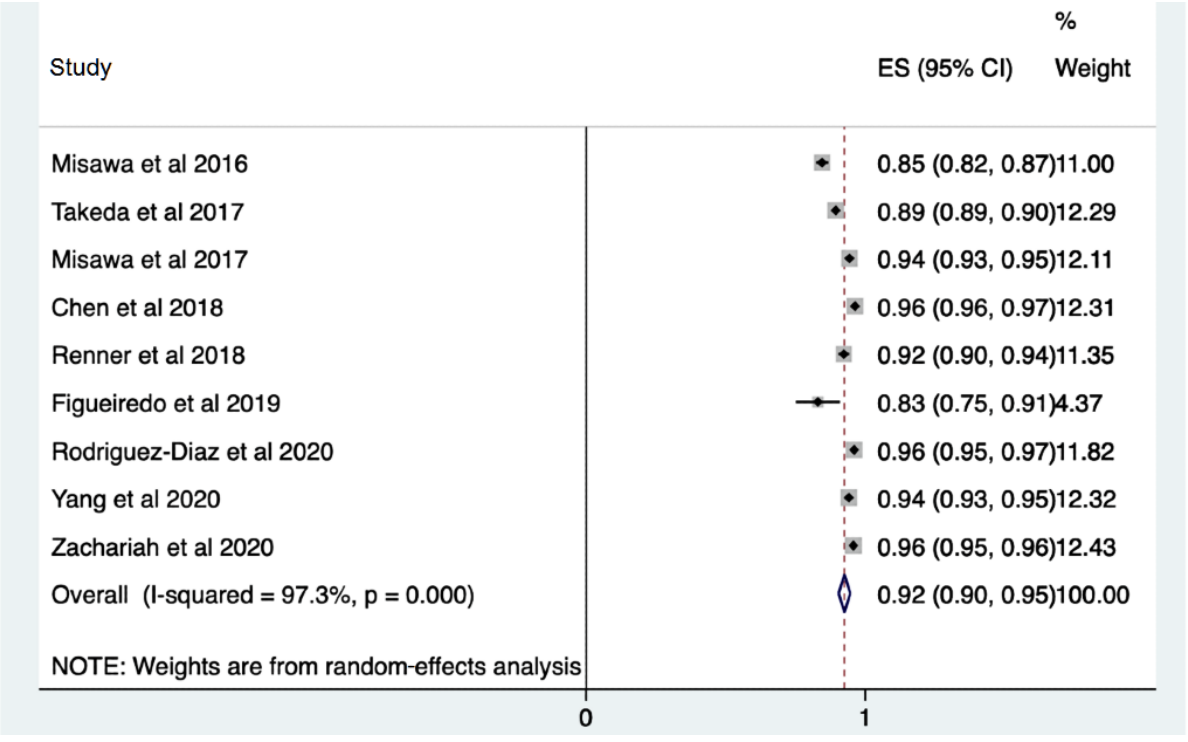
Figure 2. Pooled analysis of specificity of polyp detection by the use of lesions or polyps within video frames or images. Effect sizes (ES) are shown with 95% CIs. A random-effects model was used.



Characterization of a Detected Polyp

There were 9 studies reporting diagnostic accuracy characteristics for computer analysis of single image frames. These included a total of 22,862 images and demonstrated a sensitivity of 0.92 (95% CI 0.90-0.95; Figure 3), a specificity of 0.79 (95% CI 0.68-0.91), and an accuracy of 0.87 (95% CI 0.83-0.91). A further 20 studies assessed the diagnostic accuracy of techniques for predicting the histological diagnosis of a polyp, with a sensitivity of 0.94 (95% CI 0.92-0.95), a specificity of 0.87 (95% CI 0.83-0.90), and an accuracy of 0.91 (95% CI 0.88-0.93). A total of 16 studies analyzed diagnostic accuracy using images or video clips from a cohort of 4001 patients having undergone colonoscopy. These studies showed a sensitivity of 0.94 (95% CI 0.92-0.95), a specificity of 0.82 (95% CI 0.73-0.91), and an accuracy of 0.90 (95% CI 0.86-0.94).

Figure 3. Pooled analysis of sensitivity of polyp characterization by the use of images. Effect sizes (ES) are shown with 95% CIs. A random-effects model was used.



XSL•FO
RenderX

PDR and ADR for Polyp Detection: RCTs

The 8 RCTs consisted of a total of 5577 patients: 2438 patients in the AI group and 2463 patients in the control group with standard colonoscopy alone [56-59]. These captured data prospectively with the use of deep learning methods on real-time video streams from colonoscopy.

The meta-analysis showed a significant increase in pooled PDR in patients with the use of AI for polyp detection during colonoscopy compared with patients who had standard colonoscopy (odds ratio [OR] 1.75, 95% CI 1.56-1.96; *P*<.001; Figure 4). The PDR ranged from 38% to 64% when using AI, with a median of 45%. When comparing patients undergoing colonoscopy with the use of AI to those having standard colonoscopy, there was also a significant increase in ADR (OR 1.53, 95% CI 1.32-1.77; *P*<.001; Figure 5). The ADR ranged from 16% to 55% with a median of 34% when using AI technology compared to standard colonoscopy.

Figure 4. Pooled analysis of polyp detection rate. Odds ratios are shown with 95% CIs. A random-effects model was used for the meta-analysis.

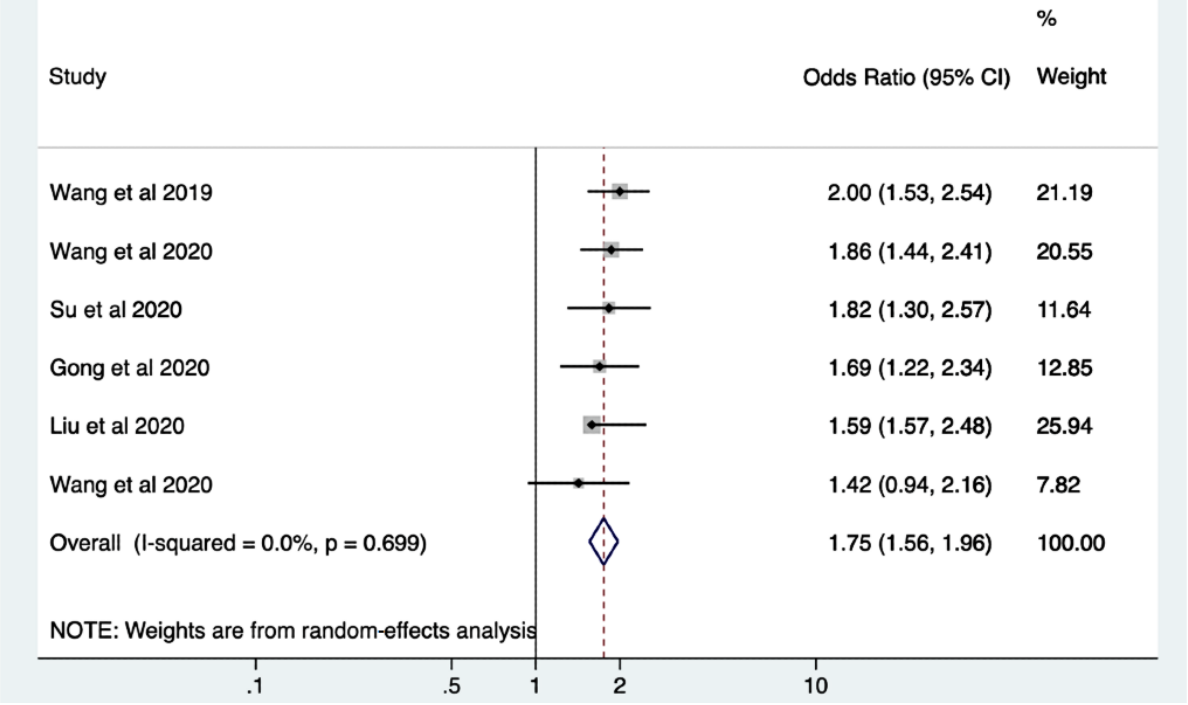
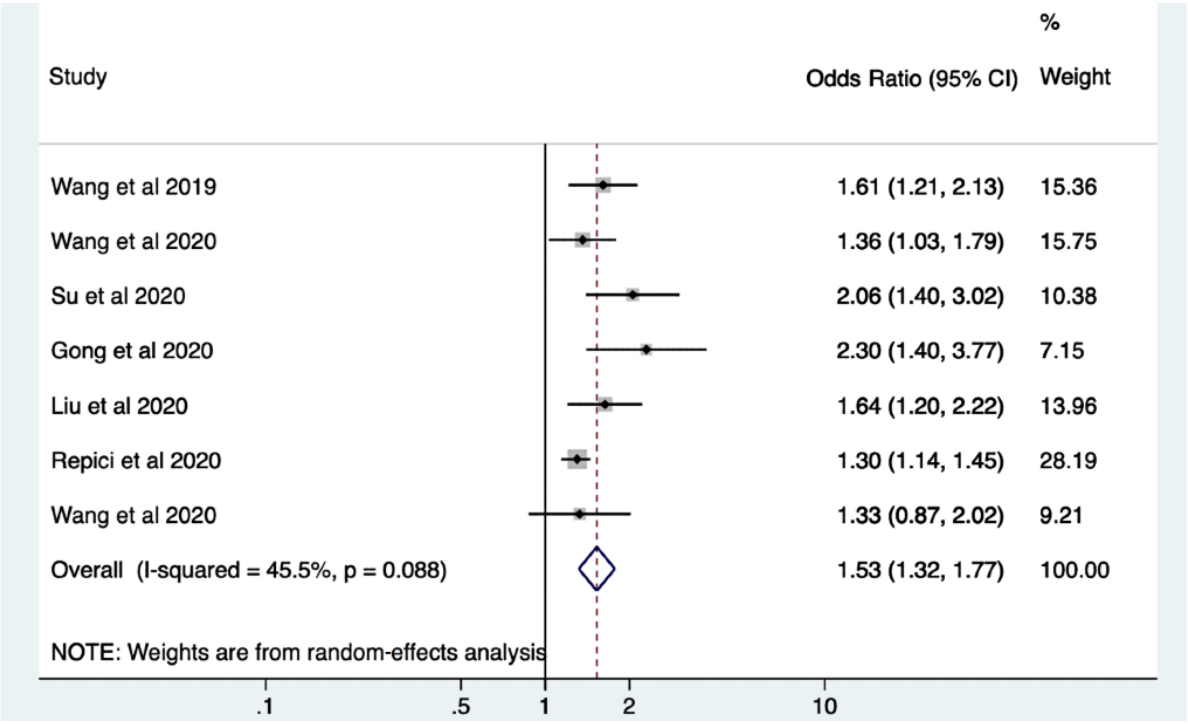


Figure 5. Pooled analysis of adenoma detection rate. Odds ratios are shown with 95% CIs. A random-effects model was used for the meta-analysis.



XSL•FO
RenderX

Heterogeneity of Studies

There was a high degree of variation between studies. The heterogeneity was statistically significant when comparing the studies for polyp detection and characterization and assessing for sensitivity, specificity, and accuracy (P<.05). The lowest variation for polyp detection was among the studies assessing accuracy with polyp data (I^2=86.3%), and the highest was among those analyzing the sensitivity of machine learning systems using image data sets (I^2=99.9%). When considering studies for polyp characterization, the heterogeneity was lowest for studies analyzing sensitivity using patient data sets (I^2=51.1%) and highest when assessing specificity using image data sets (I^2=99.9%). Within the RCTs assessed, there was found to be a low degree of heterogeneity for PDR (I^2=0%; P=.70) and a moderate degree of heterogeneity for ADR (I^2=45.5%; P=.09). These results were not statistically significant.

Quality Assessment

The assessment of bias for the studies when using the QUADAS-2 tool is depicted in Table S1 in Multimedia Appendix 2. Most of the RCTs scored 3 or more on the Jadad scale and were, therefore, considered to be of good quality (Table S2 in Multimedia Appendix 2). One study scored 2, suggesting poor quality, but after reviewing the paper and its evidence in detail, the paper was included in the final analysis [64]. This is because despite the lack of mention of blinding, the selection process for participants was justified with consecutive patients enrolled, and there were no concerns regarding applicability. The paper matched the selection criteria of our study and was otherwise in line with other studies that were included.

Discussion

Principal Findings

The aim of this systematic review and meta-analysis was to examine the current status of diagnostic accuracy for AI-based technologies in the detection and characterization of colorectal polyps. We found a wide variety of machine learning systems being used for polyp detection and characterization in numerous studies. The overall diagnostic accuracy for these systems to detect polyps was high, predominantly with sensitivities, specificities, and accuracies above 84%. When characterizing polyps, the majority of machine learning systems had sensitivities, specificities, and accuracies above 82%. These outcomes show good results for current machine learning systems and algorithms to detect and characterize polyps, and indirectly in regard to the rate of false positives.

This meta-analysis highlights a significant increase in PDR and ADR when using AI systems in conjunction with colonoscopy in real time to detect polyps in the colon and rectum with an overall OR of 1.75 (95% CI 1.56-1.96; P<.05) and 1.53 (95% CI 1.32-1.77; P<.05), respectively. The UK key performance indicators and quality assurance standards for colonoscopy dictate that the minimal ADR should be 15%, with an aspirational target of 20% [65]. It has previously been shown that endoscopists with an ADR of less than 20% had a hazard ratio for interval cancer that was 10 times higher than those with an ADR of greater than 20% [66]. All RCTs in this review were shown to have an ADR of greater than 15% when detecting polyps with the use of an AI system, the majority of which highlighted an ADR of greater than 25% [56-58]. These outcomes are a promising start for the use of AI to detect missed polyps and, thus, may lead to a reduction in CRC incidence.

The assessment of quality of the diagnostic accuracy studies included in this paper highlighted an overall low risk of bias, justifying the validity of the study results and implying that their results may be applicable to clinical practice. The main area of bias in the RCTs was in the process of blinding. This may have contributed to an overestimation in the effects of AI in polyp detection.

There are many limitations within the published studies (Table S1 in Multimedia Appendix 3). Factors contributing to the miss rate of polyps are multifactorial and include patient-related factors, polyp-related factors, and image-related factors [67,68]. It is encouraging to note that a variety of imaging modalities were used in the studies in this review, since this will improve applicability in a clinical setting. We note that most studies with image enhancement techniques have used NBI, and it will be important to validate the performance of AI systems in endoscopy using image enhancement approaches from other manufacturers (eg, i-scan from PENTAX Medical and blue laser imaging from Fujifilm Corporation). Some studies analyzing polyp characterization used magnification NBI [11,36,38,69]. This imaging modality is not commonly used in Western endoscopic practice, so is less applicable to a health care setting in the Western world. Although there has been significant development in computer-assisted technologies to increase ADR, issues with image quality still remain. Many studies in this review excluded images that were blurred or of poor quality when assessing diagnostic accuracy of the machine learning systems. [27,40,42,51]. Recent RCTs have tried to tackle this problem by developing models to recognize blurry frames [58,59]. Other studies excluded images with poor bowel preparation [27,36,48]. Adequate bowel cleansing is vital for complete mucosal inspection; however, it has been shown in a meta-analysis that low-quality preparation does not significantly affect ADR, since these patients frequently undergo repeat colonoscopy [70]. Most RCTs included in this review used the Boston Bowel Preparation Scale [71] to assess adequacy of bowel preparation.

Sufficient withdrawal time allows full mucosal inspection with careful examination of all folds and flexures, in an attempt to avoid missing any polyps. It has been shown that an increase in withdrawal time is associated with an increase in ADR [72]. This supports the use of withdrawal times as a quality indicator for screening colonoscopy. In preclinical studies, it is difficult to assess withdrawal times given the use of still images and video clips. In the RCTs assessed, the withdrawal times—excluding biopsy time—were mostly higher with the use of AI-based technology, although not significantly so in all studies (Table 3). However, the ability to record the withdrawal time is equally important [58,59]. This may suggest that quality control during colonoscopy examinations can be maintained with the use of machine learning.

XSL•FO
RenderX

Given the fact that AI is a relatively new and evolving area of medical practice, there is a lack of evidence-based standards to support its development. This is highlighted through the inconsistencies in validating the machine learning systems in each study. The data used for training the algorithms vary in type, for example, as either a static image from the colonoscopy [45,46] or an image of a polyp [21,47], and in number, with some studies having very small sample sizes [21,52]. We acknowledge the high degree of heterogeneity in the included studies, which may, in part, be explained by the wide range of approaches or algorithms used. This may suggest that our findings are applicable to a wide range of study settings and outcomes. However, the high degree of heterogeneity also emphasizes the issue of inconsistencies within the development of AI systems and, thus, weakens their design and may hinder implementation of the AI systems in a clinical setting. In order to address this problem, we are developing a new multidisciplinary, consensus-based reporting standards statement called STARD-AI (Standards for Reporting of Diagnostic Accuracy Studies–Artificial Intelligence). It is being developed to provide stringent guidelines for all AI-based clinical trials that report diagnostic accuracy [73,74].

The lack of standards among these studies introduces an element of selection bias. In traditional computer programming, intelligent systems were built by writing models by hand and, therefore, understanding the rules from which conclusions were made. Neural networks and deep learning techniques are criticized for their “black box” problem, in failing to produce an intelligible description of the results produced. This creates tension between our need for explanations and our interests in efficiency. Most studies in this systematic review did not reveal their algorithms, which begs one to question whether they only used the algorithms that were most successful in producing the desired outcome without understanding the process underlying it.

Multiple other factors contribute to the lack of applicability of these studies in clinical practice. Many of the studies about polyp detection and characterization have been carried out in Japan [46,50,51] or China [19,56,59], and differences in polyp biology and tumorigenesis may limit application to Western endoscopy practice [75]. Furthermore, for real-time detection to be successful, the operation of the AI system to detect and characterize polyps must be fast, practical, and nondisruptive to workflow. However, most current studies are designed in a nonclinical environment and carried out retrospectively, with only a handful of recent RCTs. More RCTs are needed to provide prospective data by testing the machine learning systems while a colonoscopy procedure is undertaken.

The financial implications of introducing an AI system to endoscopy should be considered. The studies in this review lack evidence to show that AI systems would be cost-effective. Before clinical application, studies must demonstrate that the current burden on health care systems and histopathology departments can be relieved, both in view of workload and in terms of costs. A very recent study examining the use of AI combined with the diagnose-and-leave strategy for diminutive polyps has found substantial reductions in the cost of colonoscopy based on prospective data [76]. This is an encouraging outcome, but more studies are needed.

The role of the health care workforce must also be considered in a time of developing AI systems. At present, real-time detection systems during colonoscopy are not able to operate independently of human direction, but understanding the change in the role of the endoscopist and nurses will be crucial for the future. In addition, a skills gap to prepare the workforce for AI will need to be addressed. The refinement of machine learning systems in detecting polyps will eventually lead to the use of AI in conjunction with all routine colonoscopy procedures. This will allow the procedure to be performed by staff who will not require the lengthy training or accreditation [77]. In this scenario, only patients with complex polyps requiring more advanced management may need to be referred to expert endoscopists.

It is important to also consider some of the ethical dilemmas that arise from the use of AI in health care. The aim of AI in polyp detection and characterization is to introduce machine learning as a “checker system” for the endoscopist. As a result, incorporation of AI into endoscopy should be encouraged as a complementary tool and not as a replacement for a clinician. For this reason, a high degree of accuracy is required from AI systems. We expect that they operate with 100% sensitivity and a low rate of false positives. However, AI is not yet free from bias or errors, and an AI decision support tool could easily succumb to automation bias when its predictions are almost always followed by the endoscopist [78]. Machine learning systems can also unintentionally reproduce or magnify existing biases of their training data sets and exacerbate health disparities [79]. Many of the studies in this meta-analysis, for example, have excluded patients with IBD or sessile serrated polyps [39,43,56], limiting their applicability for these populations. We recognize that these other cohorts of patients, including those with benign colonic pathologies and not exclusively polyps, are important to include in such research. However, this technology is still in its infancy and these patient groups represent a minority. It is difficult and not entirely feasible to create validated AI algorithms for all patient cohorts until the technology is more established and works well in its own right.

Although this systematic review has shown the performance of the AI systems to be satisfactory, the majority of the studies are preclinical trials that have not addressed these clinical needs. As a result, there remains a lack of confidence by endoscopists and patients to fully adopt the system as a whole. The clinical expectations exceed the aims of the machine learning algorithms. To fully support the incorporation of an AI system into routine practice, the diagnostic accuracy for polyp detection and characterization must meet the desired threshold, while also providing confidence that quality requirements will be fulfilled.

A further two challenges threaten the ability for AI to thrive in health care: patient confidentiality and accountability. The lack of stringent policies for the use of training data in AI means that the methods used to deidentify patient information are weak, and we suggest that standardized guidance is required for the consent of collection and use of patient data for AI training purposes. Once an algorithm-based health care system is

XSL•FO
RenderX

operational, the question of accountability arises. In the case that a machine learning system working in unison with an endoscopist detects and characterizes a polyp as hyperplastic when, in fact, it is adenomatous, who is held liable for this mistake? A robust legal framework in association with national and international endoscopy representative groups (eg, the Joint Advisory Group on Gastrointestinal Endoscopy in the United Kingdom and the ASGE in the United States) for the use of AI in health care is vital to protect endoscopists and patients. Addressing these important concerns will help build confidence and trust among patients and doctors for the use of machine learning in the delivery of care.

Conclusions

This systematic review and meta-analysis highlights the growing interest in the field of polyp detection and characterization during colonoscopy using AI. The current accuracy of machine learning for this role is high. There is potential to improve ADR and, consequently, reduce the incidence of CRC.

However, AI and machine learning systems are still evolving. Firstly, higher-quality research with modern trial designs is needed in this field, with particular attention on using larger data sets and by validating the AI systems prospectively in a clinical setting. Secondly, these systems must provide quality assurance with a robust ethical and legal framework before they can be fully embraced by clinicians and patients in the future.

Acknowledgments

Infrastructure support for this research was provided by the National Institute for Health Research Imperial Biomedical Research Centre.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for studies to include.
[DOCX File , 13 KB - jmir_v23i7e27370_app1.docx]

Multimedia Appendix 2
Quality assessment of the studies.
[DOCX File , 21 KB - jmir_v23i7e27370_app2.docx]

Multimedia Appendix 3
Limitations within the published studies.
[DOCX File , 13 KB - jmir_v23i7e27370_app3.docx]

References

1. Globocan 2018. Colorectal cancer: Number of new cases in 2018, both sexes, all ages. International Agency for Research on Cancer. 2018. URL: http://gco.iarc.fr/today [accessed 2019-11-01]
2. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012 Feb 23;366(8):687-696. [doi: 10.1056/nejmoa1100370]
3. Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: A systematic review and meta-analysis. Gastroenterology 2019 May;156(6):1661-1674.e11. [doi: 10.1053/j.gastro.2019.01.260] [Medline: 30738046]
4. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014 Apr 03;370(14):1298-1306 [FREE Full text] [doi: 10.1056/NEJMoa1309086] [Medline: 24693890]
5. Coe SG, Crook JE, Diehl NN, Wallace MB. An endoscopic quality improvement program improves detection of colorectal adenomas. Am J Gastroenterol 2013 Feb;108(2):219-226; quiz 227. [doi: 10.1038/ajg.2012.417] [Medline: 23295274]
6. Kaminski MF, Anderson J, Valori R, Kraszewska E, Rupinski M, Pachlewski J, et al. Leadership training to improve adenoma detection rate in screening colonoscopy: A randomised trial. Gut 2016 Apr;65(4):616-624 [FREE Full text] [doi: 10.1136/gutjnl-2014-307503] [Medline: 25670810]
7. Killock D. AI outperforms radiologists in mammographic screening. Nat Rev Clin Oncol 2020 Mar;17(3):134. [doi: 10.1038/s41571-020-0329-7] [Medline: 31965085]
8. Mori Y, Kudo S, Misawa M, Mori K. Simultaneous detection and characterization of diminutive polyps with the use of artificial intelligence during colonoscopy. VideoGIE 2019 Jan;4(1):7-10 [FREE Full text] [doi: 10.1016/j.vgie.2018.10.006] [Medline: 30623149]

XSL•FO
RenderX

9. Rex DK, Kahi C, O'Brien M, Levin T, Pohl H, Rastogi A, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2011 Mar;73(3):419-422. [doi: 10.1016/j.gie.2011.01.023] [Medline: 21353837]
10. Ignjatovic A, East JE, Suzuki N, Vance M, Guenther T, Saunders BP. Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard; DISCARD trial): A prospective cohort study. Lancet Oncol 2009 Dec;10(12):1171-1178. [doi: 10.1016/s1470-2045(09)70329-8]
11. Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016 Mar;83(3):643-649. [doi: 10.1016/j.gie.2015.08.004] [Medline: 26264431]
12. Rees CJ, Rajasekhar PT, Wilson A, Close H, Rutter MD, Saunders BP, et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: The Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. Gut 2017 May;66(5):887-895 [FREE Full text] [doi: 10.1136/gutjnl-2015-310584] [Medline: 27196576]
13. Ganz M, Yang X, Slabaugh G. Automatic segmentation of polyps in colonoscopic narrow-band imaging data. IEEE Trans Biomed Eng 2012 Aug;59(8):2144-2151. [doi: 10.1109/tbme.2012.2195314]
14. Chao W, Manickavasagan H, Krishna SG. Application of artificial intelligence in the detection and differentiation of colon polyps: A technical review for physicians. Diagnostics (Basel) 2019 Aug 20;9(3):99 [FREE Full text] [doi: 10.3390/diagnostics9030099] [Medline: 31434208]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
17. Jadad AR, Moore R, Carroll D, Jenkinson C, Reynolds DM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12. [doi: 10.1016/0197-2456(95)00134-4]
18. Karkanis S, Iakovidis D, Maroulis D, Karras D, Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans Inf Technol Biomed 2003 Sep;7(3):141-152. [doi: 10.1109/titb.2003.813794] [Medline: 14518727]
19. Fu JJ, Yu Y, Lin H, Chai J, Chen CC. Feature extraction and pattern classification of colorectal polyps in colonoscopic imaging. Comput Med Imaging Graph 2014 Jun;38(4):267-275. [doi: 10.1016/j.compmedimag.2013.12.009] [Medline: 24495469]
20. Wang Y, Tavanapong W, Wong J, Oh JH, de Groen PC. Polyp-Alert: Near real-time feedback during colonoscopy. Comput Methods Programs Biomed 2015 Jul;120(3):164-179. [doi: 10.1016/j.cmpb.2015.04.002] [Medline: 25952076]
21. Tajbakhsh N, Gurudu SR, Liang J. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans Med Imaging 2016 Feb;35(2):630-644. [doi: 10.1109/tmi.2015.2487997]
22. Fernández-Esparrach G, Bernal J, López-Cerón M, Córdova H, Sánchez-Montes C, Rodríguez de Miguel C, et al. Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. Endoscopy 2016 Sep;48(9):837-842. [doi: 10.1055/s-0042-108434] [Medline: 27285900]
23. Park SY, Sargent D. Colonoscopic polyp detection using convolutional neural networks. In: Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis. 2016 Presented at: SPIE Medical Imaging: Computer-Aided Diagnosis; February 27-March 3, 2016; San Diego, CA p. 978528. [doi: 10.1117/12.2217148]
24. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018 Oct;155(4):1069-1078.e8 [FREE Full text] [doi: 10.1053/j.gastro.2018.06.037] [Medline: 29928897]
25. Wang P, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nat Biomed Eng 2018 Oct;2(10):741-748. [doi: 10.1038/s41551-018-0301-3] [Medline: 31015647]
26. Misawa M, Kudo S, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial intelligence-assisted polyp detection for colonoscopy: Initial experience. Gastroenterology 2018 Jun;154(8):2027-2029.e3 [FREE Full text] [doi: 10.1053/j.gastro.2018.04.003] [Medline: 29653147]
27. Figueiredo P, Figueiredo I, Pinto L, Kumar S, Tsai Y, Mamonov A. Polyp detection with computer-aided diagnosis in white light colonoscopy: Comparison of three different methods. Endosc Int Open 2019 Feb;7(2):E209-E215 [FREE Full text] [doi: 10.1055/a-0808-4456] [Medline: 30705955]
28. Yamada M, Saito Y, Imaoka H, Saiko M, Yamada S, Kondo H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. Sci Rep 2019 Oct 08;9(1):14465 [FREE Full text] [doi: 10.1038/s41598-019-50567-5] [Medline: 31594962]

XSL•FO
RenderX

29. Becq A, Chandnani M, Bharadwaj S, Baran B, Ernest-Suarez K, Gabr M, et al. Effectiveness of a deep-learning polyp detection system in prospectively collected colonoscopy videos with variable bowel preparation quality. J Clin Gastroenterol 2020 Jul;54(6):554-557. [doi: 10.1097/MCG.0000000000001272] [Medline: 31789758]
30. Gao J, Guo Y, Sun Y, Qu G. Application of deep learning for early screening of colorectal precancerous lesions under white light endoscopy. Comput Math Methods Med 2020;2020:1-8 [FREE Full text] [doi: 10.1155/2020/8374317] [Medline: 32952602]
31. Guo Z, Nemoto D, Zhu X, Li Q, Aizawa M, Utano K, et al. Polyp detection algorithm can detect small polyps: Ex vivo reading test compared with endoscopists. Dig Endosc 2021 Jan;33(1):162-169. [doi: 10.1111/den.13670] [Medline: 32173917]
32. Lee JY, Jeong J, Song EM, Ha C, Lee HJ, Koo JE, et al. Real-time detection of colon polyps during colonoscopy using deep learning: Systematic validation with four independent datasets. Sci Rep 2020 May 20;10(1):8379 [FREE Full text] [doi: 10.1038/s41598-020-65387-1] [Medline: 32433506]
33. Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therap Adv Gastroenterol 2020;13:1-13 [FREE Full text] [doi: 10.1177/1756284820910659] [Medline: 32231710]
34. Misawa M, Kudo S, Mori Y, Hotta K, Ohtsuka K, Matsuda T, et al. Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video). Gastrointest Endosc 2021 Apr;93(4):960-967.e3. [doi: 10.1016/j.gie.2020.07.060] [Medline: 32745531]
35. Poon CCY, Jiang Y, Zhang R, Lo WWY, Cheung MSH, Yu R, et al. AI-doscopist: A real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices. NPJ Digit Med 2020;3:73 [FREE Full text] [doi: 10.1038/s41746-020-0281-z] [Medline: 32435701]
36. Tischendorf J, Gross S, Winograd R, Hecker H, Auer R, Behrens A, et al. Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study. Endoscopy 2010 Mar;42(3):203-207. [doi: 10.1055/s-0029-1243861] [Medline: 20101564]
37. Gross S, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. Gastrointest Endosc 2011 Dec;74(6):1354-1359. [doi: 10.1016/j.gie.2011.08.001] [Medline: 22000791]
38. Takemura Y, Yoshida S, Tanaka S, Kawase R, Onji K, Oka S, et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). Gastrointest Endosc 2012 Jan;75(1):179-185. [doi: 10.1016/j.gie.2011.08.051] [Medline: 22196816]
39. Mori Y, Kudo S, Wakamura K, Misawa M, Ogawa Y, Kutsukawa M, et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). Gastrointest Endosc 2015 Mar;81(3):621-629 [FREE Full text] [doi: 10.1016/j.gie.2014.09.008] [Medline: 25440671]
40. Misawa M, Kudo S, Mori Y, Nakamura H, Kataoka S, Maeda Y, et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. Gastroenterology 2016 Jun;150(7):1531-1532.e3 [FREE Full text] [doi: 10.1053/j.gastro.2016.04.004] [Medline: 27072671]
41. Mesejo P, Pizarro D, Abergel A, Rouquette O, Beorchia S, Poincloux L, et al. Computer-aided classification of gastrointestinal lesions in regular colonoscopy. IEEE Trans Med Imaging 2016 Sep;35(9):2051-2063. [doi: 10.1109/tmi.2016.2547947]
42. Mori Y, Kudo S, Chiu P, Singh R, Misawa M, Wakamura K, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: An international web-based study. Endoscopy 2016 Dec;48(12):1110-1118. [doi: 10.1055/s-0042-113609] [Medline: 27494455]
43. Takeda K, Kudo S, Mori Y, Misawa M, Kudo T, Wakamura K, et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. Endoscopy 2017 Aug;49(8):798-802. [doi: 10.1055/s-0043-105486] [Medline: 28472832]
44. Byrne MF, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019 Jan;68(1):94-100 [FREE Full text] [doi: 10.1136/gutjnl-2017-314547] [Medline: 29066576]
45. Komeda Y, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience. Oncology 2017;93 Suppl 1:30-34 [FREE Full text] [doi: 10.1159/000481227] [Medline: 29258081]
46. Misawa M, Kudo S, Mori Y, Takeda K, Maeda Y, Kataoka S, et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: Comparison with experts. Int J Comput Assist Radiol Surg 2017 May;12(5):757-766. [doi: 10.1007/s11548-017-1542-4] [Medline: 28247214]
47. Mori Y, Kudo S, Mori K. Potential of artificial intelligence-assisted colonoscopy using an endocytoscope (with video). Dig Endosc 2018 Apr;30 Suppl 1:52-53. [doi: 10.1111/den.13005] [Medline: 29658647]
48. Chen P, Lin M, Lai M, Lin J, Lu HH, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018 Feb;154(3):568-575. [doi: 10.1053/j.gastro.2017.10.010] [Medline: 29042219]
49. Renner J, Phlipsen H, Haller B, Navarro-Avila F, Saint-Hill-Febles Y, Mateus D, et al. Optical classification of neoplastic colorectal polyps - A computer-assisted approach (the COACH study). Scand J Gastroenterol 2018 Sep;53(9):1100-1106. [doi: 10.1080/00365521.2018.1501092] [Medline: 30270677]

XSL•FO
RenderX

50. Mori Y, Kudo S, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy. Ann Intern Med 2018 Aug 14;169(6):357. [doi: 10.7326/m18-0249]
51. Kudo S, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol 2020 Jul;18(8):1874-1881.e2. [doi: 10.1016/j.cgh.2019.09.009] [Medline: 31525512]
52. Figueiredo IN, Pinto L, Figueiredo PN, Tsai R. Unsupervised segmentation of colonic polyps in narrow-band imaging data based on manifold representation of images and Wasserstein distance. Biomed Signal Process Control 2019 Aug;53:101577. [doi: 10.1016/j.bspc.2019.101577]
53. Rodriguez-Diaz E, Baffy G, Lo W, Mashimo H, Vidyarthi G, Mohapatra SS, et al. Artificial intelligence-augmented visualization with real time histology mapping of colorectal polyps. Gastroenterology 2020 May;158(6):S-369. [doi: 10.1016/S0016-5085(20)31617-6]
54. Yang YJ, Cho B, Lee M, Kim JH, Lim H, Bang CS, et al. Automated classification of colorectal neoplasms in white-light colonoscopy images via deep learning. J Clin Med 2020 May 24;9(5):1593 [FREE Full text] [doi: 10.3390/jcm9051593] [Medline: 32456309]
55. Zachariah R, Samarasena J, Luba D, Duh E, Dao T, Requa J, et al. Prediction of polyp pathology using convolutional neural networks achieves "resect and discard" thresholds. Am J Gastroenterol 2020 Jan;115(1):138-144 [FREE Full text] [doi: 10.14309/ajg.0000000000000429] [Medline: 31651444]
56. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. Gut 2019 Oct;68(10):1813-1819 [FREE Full text] [doi: 10.1136/gutjnl-2018-317500] [Medline: 30814121]
57. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): A double-blind randomised study. Lancet Gastroenterol Hepatol 2020 Apr;5(4):343-351. [doi: 10.1016/s2468-1253(19)30411-x]
58. Su J, Li Z, Shao X, Ji C, Ji R, Zhou R, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: A prospective randomized controlled study (with videos). Gastrointest Endosc 2020 Feb;91(2):415-424.e4. [doi: 10.1016/j.gie.2019.08.026] [Medline: 31454493]
59. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): A randomised controlled study. Lancet Gastroenterol Hepatol 2020 Apr;5(4):352-361. [doi: 10.1016/s2468-1253(19)30413-3]
60. Liu W, Zhang Y, Bian X, Wang L, Yang Q, Zhang X, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26(1):13. [doi: 10.4103/sjg.sjg_377_19]
61. Luo Y, Zhang Y, Liu M, Lai Y, Liu P, Wang Z, et al. Artificial intelligence-assisted colonoscopy for detection of colon polyps: A prospective, randomized cohort study. J Gastrointest Surg 2020 Sep 23:1-8. [doi: 10.1007/s11605-020-04802-4] [Medline: 32968933]
62. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020 Aug;159(2):512-520.e7. [doi: 10.1053/j.gastro.2020.04.062] [Medline: 32371116]
63. Wang P, Liu P, Glissen Brown JR, Berzin TM, Zhou G, Lei S, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. Gastroenterology 2020 Oct;159(4):1252-1261.e5. [doi: 10.1053/j.gastro.2020.06.023] [Medline: 32562721]
64. Liu W, Zhang Y, Bian X, Wang L, Yang Q, Zhang X, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26(1):13. [doi: 10.4103/sjg.sjg_377_19]
65. Rees CJ, Thomas Gibson S, Rutter MD, Baragwanath P, Pullan R, Feeney M, British Society of Gastroenterology, the Joint Advisory Group on GI Endoscopy, the Association of Coloproctology of Great Britain and Ireland. UK key performance indicators and quality assurance standards for colonoscopy. Gut 2016 Dec;65(12):1923-1929 [FREE Full text] [doi: 10.1136/gutjnl-2016-312044] [Medline: 27531829]
66. Kaminski MF, Regula J, Kraszewska E, Polkowski M, Wojciechowska U, Didkowska J, et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010 May 13;362(19):1795-1803. [doi: 10.1056/nejmoa0907667]
67. Kim NH, Jung YS, Jeong WS, Yang H, Park S, Choi K, et al. Miss rate of colorectal neoplastic polyps and risk factors for missed polyps in consecutive colonoscopies. Intest Res 2017 Jul;15(3):411-418 [FREE Full text] [doi: 10.5217/ir.2017.15.3.411] [Medline: 28670239]
68. Leufkens A, van Oijen M, Vleggaar F, Siersema P. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy 2012 May;44(5):470-475. [doi: 10.1055/s-0031-1291666] [Medline: 22441756]
69. Gross S, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. Gastrointest Endosc 2011 Dec;74(6):1354-1359. [doi: 10.1016/j.gie.2011.08.001] [Medline: 22000791]
70. Clark BT, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: Systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. Am J Gastroenterol 2014 Nov;109(11):1714-1723; quiz 1724 [FREE Full text] [doi: 10.1038/ajg.2014.232] [Medline: 25135006]

XSL•FO
RenderX

71. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: A valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009 Mar;69(3 Pt 2):620-625 [FREE Full text] [doi: 10.1016/j.gie.2008.05.057] [Medline: 19136102]
72. Shaukat A, Rector TS, Church TR, Lederle FA, Kim AS, Rank JM, et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. Gastroenterology 2015 Oct;149(4):e14-e15. [doi: 10.1053/j.gastro.2015.08.028]
73. Reporting guidelines under development for other study designs. The EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network. URL: https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/ [accessed 2020-03-24]
74. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med 2020 Jun;26(6):807-808. [doi: 10.1038/s41591-020-0941-1] [Medline: 32514173]
75. Hur J, Baek MJ. Limitation and value of using the adenoma detection rate for colonoscopy quality assurance. Ann Coloproctol 2017 Jun;33(3):81 [FREE Full text] [doi: 10.3393/ac.2017.33.3.81] [Medline: 28761865]
76. Mori Y, Kudo S, East JE, Rastogi A, Bretthauer M, Misawa M, et al. Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: An add-on analysis of a clinical trial (with video). Gastrointest Endosc 2020 Oct;92(4):905-911.e1. [doi: 10.1016/j.gie.2020.03.3759] [Medline: 32240683]
77. Topol E. Artificial intelligence and robotics. The Topol Review. Preparing the Healthcare Workforce to Deliver the Digital Future. London, UK: NHS England; 2019 Feb. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2021-07-01]
78. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc 2020 Dec 09;27(12):2020-2023 [FREE Full text] [doi: 10.1093/jamia/ocaa094] [Medline: 32574353]
79. Rigby MJ. Ethical dimensions of using artificial intelligence in health care. AMA J Ethics 2019 Feb 01;21:121-124 [FREE Full text] [doi: 10.1001/amajethics.2019.121]

Abbreviations

ADR: adenoma detection rate
AI: artificial intelligence
ASGE: American Society of Gastrointestinal Endoscopy
CAD: computer-aided diagnosis
CNN: convolutional neural network
CRC: colorectal cancer
DCNN: deep convolutional neural network
IBD: inflammatory bowel disease
NBI: narrow band imaging
OR: odds ratio
PDR: polyp detection rate
PIVI: Preservation and Incorporation of Valuable endoscopic Innovations
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2
RCT: randomized controlled trial
STARD-AI: Standards for Reporting of Diagnostic Accuracy Studies–Artificial Intelligence

Edited by R Kukafka; submitted 22.01.21; peer-reviewed by K Lam, F Iqbal; comments to author 27.01.21; revised version received 09.03.21; accepted 06.05.21; published 14.07.21.

Please cite as:
Nazarian S, Glover B, Ashrafian H, Darzi A, Teare J
Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(7):e27370
URL: https://www.jmir.org/2021/7/e27370
doi:10.2196/27370
PMID:34259645

XSL•FO
RenderX

©Scarlet Nazarian, Ben Glover, Hutan Ashrafian, Ara Darzi, Julian Teare. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Implementation of Telemental Health Services Before COVID-19: Rapid Umbrella Review of Systematic Reviews

Phoebe Barnett[1,2], MSc; Lucy Goulding[3], PhD; Cecilia Casetta[4], MD; Harriet Jordan[4,5], MSc; Luke Sheridan-Rains[2], PhD; Thomas Steare[2], MSc; Julie Williams[6], PhD; Lisa Wood[7], PhD; Fiona Gaughran[4,5*], MD; Sonia Johnson[2,8*], DM

[1]Centre for Outcomes Research and Effectiveness, Division of Psychology and Language Sciences, University College London, London, United Kingdom

[2]NIHR Mental Health Policy Research Unit, Division of Psychiatry, University College London, London, United Kingdom

[3]King's Improvement Science, Centre for Implementation Science, King's College London, London, United Kingdom

[4]NIHR Applied Research Collaboration, King's College London, London, United Kingdom

[5]Institute of Psychiatry, Psychology and Neuroscience, King's College London and South London and Maudsley NHS Trust, London, United Kingdom

[6]Centre for Implementation Science, Health Service and Population Research Department, King's College London, London, United Kingdom

[7]Division of Psychiatry, University College London, London, United Kingdom

[8]Camden and Islington NHS Foundation Trust, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Phoebe Barnett, MSc
Centre for Outcomes Research and Effectiveness
Division of Psychology and Language Sciences
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 020 7612 6143
Email: phoebe.barnett@ucl.ac.uk

Abstract

Background: Telemental health care has been rapidly adopted for maintaining services during the COVID-19 pandemic, and a substantial interest is now being devoted in its future role. Service planning and policy making for recovery from the pandemic and beyond should draw on both COVID-19 experiences and the substantial research evidence accumulated before this pandemic.

Objective: We aim to conduct an umbrella review of systematic reviews available on the literature and evidence-based guidance on telemental health, including both qualitative and quantitative literature.

Methods: Three databases were searched between January 2010 and August 2020 for systematic reviews meeting the predefined criteria. The retrieved reviews were independently screened, and those meeting the inclusion criteria were synthesized and assessed for risk of bias. Narrative synthesis was used to report these findings.

Results: In total, 19 systematic reviews met the inclusion criteria. A total of 15 reviews examined clinical effectiveness, 8 reported on the aspects of telemental health implementation, 10 reported on acceptability to service users and clinicians, 2 reported on cost-effectiveness, and 1 reported on guidance. Most reviews were assessed to be of low quality. The findings suggested that video-based communication could be as effective and acceptable as face-to-face formats, at least in the short term. Evidence on the extent of digital exclusion and how it can be overcome and that on some significant contexts, such as children and young people's services and inpatient settings, was found to be lacking.

Conclusions: This umbrella review suggests that telemental health has the potential to be an effective and acceptable form of service delivery. However, we found limited evidence on the impact of its large-scale implementation across catchment areas. Combining previous evidence and COVID-19 experiences may allow realistic planning for the future implementation of telemental health.

(J Med Internet Res 2021;23(7):e26492) doi:10.2196/26492

XSL•FO
RenderX

KEYWORDS

umbrella review; remote; telemental health; telepsychiatry; video-based therapy; COVID-19; remote therapy; review; mental health; therapy; eHealth; telemedicine; mobile phone

Introduction

Background

The use of technologies such as phone or video calls between mental health professionals and other mental health professionals, patients, service users, family members, or carers to deliver mental health care (telemental health) has become an essential tool in recent months; importantly, it has taken a central role internationally in maintaining mental health services during the COVID-19 pandemic [1]. Policy makers and mental health professionals, along with mental health service users, now express interest in continuing some use of these technologies long term, even in the absence of pandemic-related social distancing requirements [1-3]. The potential benefits of remote technologies extend beyond adaptation to government social distancing guidelines, allowing the efficiency and flexibility of mental health services to be maximized. The mobilization of telemental health during the pandemic has happened largely ad hoc, thereby achieving remarkably rapid but highly variable implementation. This emergency response has mostly occurred without systematic references to the literature. To plan the effective and acceptable deployment of telemental health beyond the pandemic, it is crucial that we now review all the relevant evidence regarding potential impacts, challenges, and outcomes of widespread remote technology utilization and identify the key mechanisms for its acceptable integration into routine care [4].

Telemental health offers a number of potential benefits that generate a significant interest among service providers not only during the pandemic but also in the long term. For service users across a range of populations, settings, and conditions [5], potential benefits include convenience and improved accessibility, particularly where issues such as physical mobility difficulties, anxiety, or paranoia impede a face-to-face contact [1]. Potential advantages for staff include reduced environmental impact, greater convenience with opportunities for working at home, and ease of communication within and between mental health teams [2]. Although some have argued that problems with building rapport [6] along with privacy and safety concerns may hinder the implementation of remote care, service users have been found to report such apprehensions less than clinicians [7]. Several studies have also suggested that telemental health may be more cost-effective than face-to-face delivery [7].

Despite potential benefits and a substantial body of relevant research, the delivery of remote working remained very limited in most countries before the pandemic with substantial implementation barriers [8], along with the potential for inequalities to be exacerbated. Digital exclusion is an important concern for service users who lack the necessary skills, space, equipment, and monetary resources to access web-based treatment. This is more marked in marginalized groups such as people from BAME (Black, Asian, and minority ethnic) and low socioeconomic status backgrounds [9]. Further risks include the loss of privacy and deterioration in therapeutic relationships [1,10-12]. Staff participation is also impeded by technological and environmental difficulties, and they express reservations regarding the quality of assessments, deterioration of therapeutic relationships, and limitations in the extent to which there is a focus on physical as well as mental health [8,11,12].

Objectives

Therefore, the potential benefits and disadvantages of telemental health are finely balanced. Risks of longer-term rollout of telemental health without close attention to intended and unintended consequences include the digital exclusion of some of those already most disadvantaged and decline in the quality of care and, potentially, of outcomes. One source with the potential to inform policy makers and service planners in their future telemental health strategies is the substantial body of research studies published before the pandemic. This paper therefore aims to provide a rapid summary of the existing literature on the effectiveness, cost-effectiveness, barriers and facilitators for implementation, acceptability, and reach of remote interventions for the assessment and treatment of mental health problems. Our objective is to identify, appraise, and synthesize the systematic reviews of literature and guidance on telemental health, including qualitative and quantitative outcomes using the *umbrella review* or *review of reviews* methodology. Umbrella reviews are useful in summarizing a broad evidence base to inform policy [13]. It is hoped that the results will help illuminate the benefits and remaining challenges while implementing telehealth technologies during the remainder of the pandemic and in the perhaps permanently changed reality that follows.

Methods

Overview

A rapid umbrella review was conducted guided by the World Health Organization practical guide for Rapid Reviews to Strengthen Health Policy and Systems [14] and adhering to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15] and umbrella review guidance [16]. In line with the agreed rapid review methodology, this review aims to provide a timely but robust answer to the research question by accelerating some aspects of the systematic review process while maintaining transparency and protocol-driven decision making throughout [14]. The protocol was prospectively registered on PROSPERO (CRD42020208085).

Search Strategy and Selection Criteria

Overview

The search strategy implemented a combination of keywords and subject heading searches across PsycINFO (01/01/2010-26/08/2020), PubMed (01/01/2010-26/08/2020), and the Cochrane Database of Systematic Reviews

XSL•FO
RenderX

(01/01/2010-26/08/2020). Searches combined terms for systematic reviews, mental health disorders, and telemental health. The full search strategy is available in Multimedia Appendix 1. We included systematic reviews that met the following criteria.

Population

Staff working within the field of mental health, people receiving mental health care or with mental health diagnoses, family members, or carers of people receiving mental health care were included. In line with our focus on people whose conditions tend to make them eligible for telemental health care, we also included people with dementia, neurodevelopmental disorders, and addiction but excluded people with primary sleep disorders, unless combined with other mental health problems.

Interventions

Any form of spoken or written communication conducted between mental health professionals and patients, service users, family members, carers, or other mental health professionals using either the internet or the telephone were included. We excluded the reviews of digital interventions where the primary aim of the technology was not to facilitate direct therapeutic contact with a mental health professional; thus, for example, we excluded apps and websites delivering assessment or treatment in a digital format.

Outcomes

Reviews reporting at least one of the following: implementation outcomes (outcomes relating to the process of care, adherence to intended models, uptake and coverage, and barriers and facilitators to implementation), acceptability outcomes (including staff and service user satisfaction, and experiences of the therapeutic relationship and communication), clinical effectiveness, cost-effectiveness, or evidence-based guidance for telemental health were included. Qualitative and quantitative data were analyzed.

Design

Systematic reviews with or without meta-analyses, realist reviews, and qualitative meta-syntheses were included. We considered reviews to be of sufficient quality for inclusion if they searched at least three databases based on recommendations by Cochrane and the Assessment of Multiple Systematic Reviews (AMSTAR) [17], meaning that at least two databases plus subject-specific databases were searched. In line with recommendations for conducting systematic reviews for quantitative data [15], quantitative reviews were also required to include appraisal of the quality of included studies.

Owing to the rapid nature of the review, we limited our search to reviews published since January 2010 and those available in the English language. This was a pragmatic decision because studies published before 2010 would still be picked up within systematic reviews.

A total of 3 reviewers (PB, LG, and CC) double screened 10% of titles and abstracts, and disagreements were discussed until consensus was reached. The remaining titles were independently screened, and studies not meeting the inclusion criteria were excluded. Full-text articles were subsequently reviewed by 5 reviewers (PB, TS, LG, CC, and LW). A selection of full texts was double checked to ensure consistency, and any reviews that did not facilitate a straightforward inclusion or exclusion decision were discussed with the wider review group.

Data Extraction

A total of 7 reviewers (LG, CC, PB, TS, LSR, JW, and HIJ) extracted data from the included reviews using a Microsoft Excel-based form. In total, 10% of extractions were double checked by a second reviewer, and inconsistencies were discussed and corrected. The extracted data included citation details, objectives, type of review, participant details (including gender, ethnicity, age, and mental health diagnosis accompanied by staff details where relevant), types of telemental health intervention reviewed, setting and context (mental health service, community, inpatient or residential, or primary mental health care service), number of databases sourced and searched, date range of database searching, the publication date range of studies included in the review informing each outcome of interest, number of included studies, types of studies and country of origin of studies included, instrument used to appraise the primary studies and the rating of their quality; reported clinical, cost-effectiveness, and implementation outcomes; method of synthesis or analysis employed to synthesize the evidence; and conclusions of the review authors.

Quality Assessment

The quality of each included systematic review was assessed using the AMSTAR 2 checklist [17]. This is a revised version of the original AMSTAR checklist [18], which is a validated measure of systematic review quality [19]. The checklist was used to give each review an overall rating of quality ranging from high (0 or 1 noncritical weakness: providing an accurate and comprehensive summary of the results of the available studies that address the question of interest) to critically low (more than 1 critical flaw with or without noncritical weaknesses: the review should not be relied upon to provide an accurate and comprehensive summary of the available studies) [17]. The study quality was assessed along with data extraction. Table 1 presents the quality ratings.

XSL•FO
RenderX

Table 1. Study characteristics.

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Hareri-mana et al (2019) [20]	Mobile apps (NR[b]); smart technologies (NR); teleconferencing systems (NR); internet-based therapies (NR); Skype (videoconferencing) calls (NR)	Waiting list or TAU[c] (NR); no comparator (NR)	1946-2017	9	Pilot RCT[d] (n=2); RCT (n=2); program case analysis (n=1); quasiexperimental study (n=1); prospective design (n=1); cross-sectional survey (n=1); case study (n=1)	2032 (NR)	Depression or self-reported depressive symptoms (n=9)	>65 (NR)	NR	United States (n=5); Australia (n=1); Canada (n=1); China (n=1); The Netherlands (n=1)	Low
Dorstyn et al (2013) [21]	Telecounseling, that is, telephone, videophone, and computer (NR) or web-based digital media, that is, email, audio-only, or audio-video communication via the internet (NR)	TAU (n=3); face-to-face (n=1); minimal support or waitlist (n=2); no comparator (n=2)	1970-2013	9 (8 different samples)	RCT (n=7); single arm (n=1); nonrandomized controlled trial (n=1)	498 (n=66)	Depression or psychiatric comorbidities with depression symptoms (n=9)	54 (NR)	Hispanic, 243 (52); Latino, 139 (30); Asian, 105 (21); African American, 11 (2)	United States (n=6); Canada (n=1); Australia (n=1)	Critically low

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Berryhill et al (2019a) [22]	Video-based CBT[e] (n=12); video-based behavioral activation (n=5); video-based acceptance and behavioral therapy (n=1); video-based exposure (n=3); video-based metacognitive therapy (n=1); video-based problem-solving therapy (n=2); video-based therapy in multiple modalities (n=9)	Face-to-face psychotherapy (n=16); face-to-face or telephone (n=2); no control (n=15)	1991-2017	33	RCT (n=14); quasiexperimental (n=4); single cohort study—before and after (n=9); case study (n=4); multiple baseline design (n=1); single case interrupted time series (n=1)	NR	Depression (n=9); PTSD[f] (n=12); depression with comorbid anxiety or PTSD (n=12)	NR (10.3-80.4)	NR	NR	Critically low
Berryhill et al (2019b) [23]	Video-based CBT (n=12); video-based behavioral activation (n=3); video-based ACT[g] (n=1); video-based exposure therapy (n=2); video-based problem-solving therapy (n=1); video-based metacognitive therapy (n=1); multiple modality (n=1)	Face-to-face psychotherapy (n=20); no control (n=1)	1991-2017	21	RCT (n=6); quasiexperimental (n=4); uncontrolled (n=11)	NR	Depression (n=2); PTSD (n=7); anxiety disorder (ie, PD[h], GAD[i], and social phobia; n=5); depression or mood disorder (n=7)	NR (8-62)	NR	United States (n=10); Australia (n=6); Canada (n=5)	Critically low
Bolton and Dorstyn (2015) [24]	Internet-based CBT with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no comparator (n=4)	1970-2014	11	RCT (n=4); nonrandomized (n=7)	472 (NR)	PTSD (n=11)	40 (18-68)	NR	United States (n=6); Australia (n=3); Canada (n=1); United Kingdom (n=1)	Critically low

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Christensen et al (2019) [25]	Video consultations and telepsychiatry (n=21)	Face-to-face (n=11); no control (n=10)	2000-2017	21	RCT (n=7); surveys (n=3); intervention study (n=6); evaluation using qualitative and quantitative methods (n=1); qualitative studies (n=4)	2525 (NR)	Depression (n=6); various diagnoses (n=15)	NR	NR	United States (n=12); Canada (n=5); Spain (n=1); Australia (n=1); Hong Kong (n=1); Germany (n=1)	Low
Coughtrey and Pistrang (2018) [26]	CBT (n=12); ERPT[j] (n=1); behavioral therapy (n=1)	Face-to-face exposure response therapy (n=1); telephone emotion-focused therapy (n=1); TAU (n=5); waitlist (n=3); no comparator (n=4)	1991-2016	14	RCT (n=9); uncontrolled design (n=3); quasiexperimental (n=2)	750 (NR)	Depression (n=10; 5 with physical comorbidities); OCD[k] (n=2); anxiety disorders (n=2)	NR (32-66)	NR	United States (n=11); United Kingdom (n=2); Canada (n=1)	Low
Drag et al (2016) [27]	Videoconference (n=24)	Face-to-face (n=23); no comparator (n=1)	2000-2015	26	RCT (n=26)	Analysis of assessment, 765 (NR); analysis of efficacy 2097 (NR)	Analysis of assessment: multiple diagnoses (n=6); Alzheimer disease (n=2); schizophrenia (n=3); autism (n=1); analysis of efficacy: multiple diagnoses (n=2); PTSD (n=3); ADHD[l] (n=1); major depression (n=6); Alzheimer disease (n=1); eating disorders (n=1)	Analysis of assessment, NR (9-68); analysis of efficacy, NR (9-65)	NR	United States (n=17); Canada (n=2); Japan (n=2); China (n=1); New Zealand (n=1); India (n=1); Norway (n=1); Spain (n=1)	Low

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Garcia-Lizana and Munoiz-Mayorga (2010) [28]	Videoconference (n=10)	NR	1997-2008	11	RCT (n=10)	1054 (NR)	Multiple diseases (n=4); depression (n=2); panic disorder (n=1); PTSD (n=1); bulimia (n=1); schizophrenia (n=1)	NR	NR	United States (n=6); Canada (n=4); Spain (n=1)	Critically low
Hassan and Sharif (2019) [29]	Not specified videoconferencing treatment intervention (n=2); video-based CBT (n=7); video-based psychoeducation (n=2); video-based relapse prevention (n=1); video-based treatment management (n=1); video-based evaluation of competency to stand trial (n=1)	Face-to-face (n=14)	2000-2017	14	RCT (n=14)	1714 (NR)	Multiple (n=4); depression (n=5); panic disorder (n=1); PTSD (n=1); schizophrenia (n=1); bulimia nervosa (n=1); mental incompetency (n=1)	NR	NR	Canada (n=5); United States (n=8); Spain (n=1)	Critically low
Lin et al (2019) [30]	Psychotherapy (n=10); medication (n=3)	Face-to-face psychotherapy (n=7); telephone (n=2); TAU (n=1); no comparator (n=3)	1998-2018	13	RCT (n=7); quasiexperimental (n=1); nonrandomized pilot studies (n=2); retrospective studies (n=3)	5546 (NR–substantial variability in gender reported)	Substance use disorders, including alcohol (n=5), nicotine (n=3), opioid (n=5)	NR (30.5-52; 1 study did not report)	NR	United States (n=10); Canada (n=2); Denmark (n=1)	Moderate
Lins et al (2014) [31]	Telephone counseling (n=9)	Friendly calls (n=3); TAU (n=6)	2000-2008	12	RCT (efficacy: n=9); qualitative study (experience of intervention: n=3)	NR	Depressive symptoms (n=8); anxiety symptoms (n=1)	NR (60-66)	NR	United States (n=8); Germany (n=1); Canada or United States (n=3)	Moderate

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Muskens et al (2014) [32]	Telephone diagnostic interviewing (n=16)	Traditional face-to-face diagnostic interviewing	NR (search was conducted in 2012)	16	NR	1001 (NR)	Studies conducted diagnostic interviewing for a range of diagnoses including depression, anxiety, substance misuse, psychotic disorders, autism, PTSD, manic episodes or mania, panic disorder, social phobia, simple phobia, dysthymia. Included studies interviewed for between 1 and 21 disorders	NR (8.92-76.9)	NR	United States (n=10); United Kingdom (n=2); Brazil (n=1); Australia (n=1); Canada (n=1); Iran (n=1)	Moderate
Naslund et al (2020) [33]	Videoconference for psychiatric or neurological assessment or treatment (n=23); videotaping psychiatric histories (n=1); sending clinical information electronically to psychiatrist for diagnosis and treatment plan (n=1); therapy via text messages (n=1)	Face-to-face (n=26)	2000-2018	26	RCT (n=11); observational study (n=10); pre-post study (n=3); quasiexperimental (n=2)	17,967 (NR)	Depression (n=7); general mental disorders (n=7); child mental health (n=4); geriatric mental health (n=4); PTSD (n=2); suicidal ideation (n=1); epilepsy (n=1)	NR	NR	Canada (n=4); Colombia (n=1); United States (n=15); Spain (n=1); Germany (n=1); Australia (n=2); Israel (n=1); Hong Kong (n=1)	Critically low

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Norwood et al (2018) [34]	Video-based CBT (n=10)	Face-to-face CBT (n=10)	NR (search took place in 2018)	10	RCT (n=4); non-RCT (n=2); case studies or series (n=3); uncontrolled trial (n=1)	343 (NR)	Depression, anxiety, or mood disorder (n=3); bulimia nervosa or EDNOS[m] (n=1); PTSD (n=2); OCD (n=1); panic disorder with agoraphobia (n=1); social anxiety (n=1); NR (n=1)	NR	NR	United States (n=6); Canada (n=1); France (n=1); United Kingdom (n=1); Australia (n=1)	Moderate
Olthuis et al (2016a) [35]	Internet CBT with therapist email or telephone support (n=37); internet behavioral therapy with exposure (n=1)	Waitlist or attentional control (n=20); face to face (n=7); other internet therapies (n=6; multiple control groups (n=5)	Up to 2015	30	RCT	218 (67.1)	Social phobia (n=11); PD with or without agoraphobia (n=8); GAD (n=5); PTSD (n=2); OCD (n=2); specific phobia (n=2); mixed anxiety (n=8)	37.3 (NR)	NR	Sweden (n=18); Australia (n=14); Switzerland (n=3); Netherlands (n=2); United States (n=1)	Moderate
Olthuis et al (2016b) [36]	Internet CBT (with therapist contact) or CBT by phone (n=19)	Face-to-face (n=8); internet-based supportive counseling (n=1); TAU (n=2); telephone (n=1); self-help iCBT[n] (n=1); waiting list (n=6)	Up to 2016	19	RCT	1491 (67.7)	PTSD (n=13); subclinical PTSD (n=6)	NR	NR	United States (n=13); Sweden (n=3); Germany (n=1); Australia (n=2)	Moderate
Sansom-Daly et al (2016) [37]	N/A[o] (systematic review of guidelines)	N/A	2004-2014	20	N/A	N/A	N/A	N/A	N/A	United States (n=10); Canada (n=5); Australia (n=1); United Kingdom (n=1); Europe (n=1); South Africa (n=1); New Zealand (n=1)	Low

XSL•FO
RenderX

Study	Intervention type (number of studies)	Comparator (number of studies)	Search dates	Number of studies included	Study designs (number of studies)	Patients included, n (% female)	Diagnoses (number of studies)	Population age in years, mean (range)	Ethnicity, n (%)	Countries covered (number of studies)	Quality appraisal rating (AMSTAR 2[a])
Turgoose et al (2018) [38]	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	Up to 2018	41	NR. A mix of experimental and nonexperimental designs	4130 (NR)	PTSD (n=41)	NR	NR	United States (n=40); Canada (n=1)	Critically low

[a]AMSTAR 2: Assessment of Multiple Systematic Reviews.

[b]NR: not reported.

[c]TAU: treatment as usual.

[d]RCT: randomized controlled trial.

[e]CBT: cognitive behavioral therapy.

[f]PTSD: posttraumatic stress disorder.

[g]ACT: acceptance and commitment therapy.

[h]PD: Parkinson disease.

[i]GAD: generalized anxiety disorder.

[j]ERPT: exposure response prevention therapy.

[k]OCD: obsessive-compulsive disorder.

[l]ADHD: attention-deficit/hyperactivity disorder.

[m]EDNOS: eating disorders not otherwise specified.

[n]iCBT: internet-based cognitive behavioral therapy.

[o]N/A: not applicable.

Data Synthesis

Heterogeneity in study populations and interventions included in the review, as well as broad inclusion criteria for review study design (eg, qualitative and quantitative), prevented the quantitative pooling of syntheses. As a result, we conducted a narrative synthesis of all interventions and outcomes [39]. This allowed for a more in-depth consideration of all outcome measures and variations in the remote intervention delivery. We grouped reviews by the included population (mental health diagnosis) and further considered the variation in interventions on offer within these subgroups. This was performed for each outcome of interest. Most reviews provided a synthesis of multiple intervention types or failed to adequately differentiate them, making a more thorough comparison across formats impossible.

Results

Overview

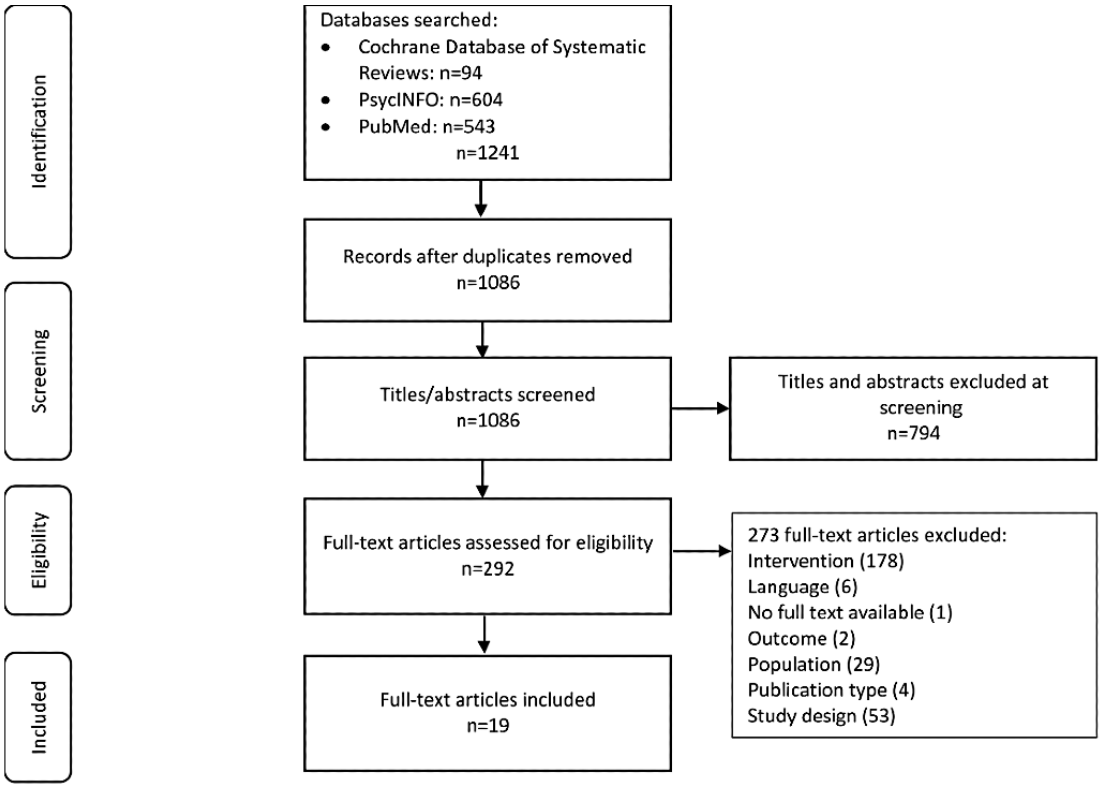
The search returned 1086 reviews, of which 292 potentially relevant full-text articles were identified. Following full-text checks, 19 reviews met the inclusion criteria (Figure 1), reporting 239 individual studies and 20 guidance documents. In total, 15 of the included reviews examined the clinical effectiveness of telemental health as compared with (1)

XSL•FO
RenderX

face-to-face interventions and assessments (n=4), (2) treatment as usual (n=2), or (3) a variety of comparators, including face-to-face, telephone, and treatment as usual (n=9). Eight reviews reported on implementation (broadly defined), including process variables, fidelity, and uptake of interventions, and 10 reviews reported outcomes related to acceptability, including the satisfaction of both service users and clinicians. In total, 1 review focused specifically on the differences in therapeutic alliances between treatment modalities. A total of 2 reviews reported cost-effectiveness, one on this topic only and the other in combination with clinical effectiveness. One review synthesized international guidance on the conduct of videoconferencing-based mental health treatments. Table 1 presents full details of the included reviews, and Figure 1 shows the information on the search and screening process.

Some primary studies were included in more than 1 review: 26 studies appeared in 2 reviews and 27 studies appeared in 3 or more. The remaining 186 studies were included in only 1 review. The double counting of primary studies due to inclusion in multiple reviews contributing to the same outcome was only found for clinical effectiveness outcomes. However, conclusions were similar across reviews, and no review had all the same studies contributing to any synthesis. Multimedia Appendix 2 [20-36,38,40-276] and Multimedia Appendix 3 [40-276] present further details of the study overlap.

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



Quality of Included Reviews

Most reviews elicited low confidence in quality appraisal because of multiple weaknesses in the study design. The most common weaknesses were a lack of explicit statements that a protocol was developed before the commencement of the review (explicit statements were reported in two reviews [20,35]), lack of duplicate study selection (duplicate selection was reported in five reviews [28,31,32,35,37]), no report of excluded studies and reasons for exclusion (exclusions were reported in two reviews [31,35]), and no report of sources of funding (sources of funding were reported in three reviews [26,27,31]). Meta-analysis was not performed in the majority of reviews, usually because of heterogeneous data or aims centering around more narrative conclusions such as acceptability (n=12) but in those that included meta-analysis [24,27,31,34-36], all except two [31,34] reviews assessed publication bias. The potential impact of risk of bias was only assessed in two reviews performing meta-analysis [31,34], but all reviews performing meta-analysis used appropriate statistical methods for combining results. The reviews eliciting higher confidence (moderate) were two Cochrane reviews [31,36]. Table 1 presents the quality ratings of reviews, and Multimedia Appendix 4 [20-32,34-38] shows the full details of quality assessments.

Clinical Outcomes

Clinical outcomes were reported in 15 reviews [20-24,26-31,34-36,38]. Across all patient populations, including patients with anxiety (n=3), posttraumatic stress disorder (PTSD; n=2), depression (n=4) (including in ethnic minorities, n=1 [21], and older adults, n=1 [20]), substance use disorders (n=1), and multiple disorders (n=4), videoconferencing interventions have been reported to cause significant reductions in symptom severity, with outcomes comparable with face-to-face controls where these were included. Telephone-based interventions tended to report similar significant reductions in symptom severity. However, a review of telephone interventions with older adults with depression [20] reported more mixed findings: reductions were reported in emergency room and hospital visits in one study, and in depression in another; however, a third

XSL•FO
RenderX

study suggested that telephone interventions did not add to benefit gained from a web-only intervention. Follow-up treatment gains were less widely reported, and conclusions were mixed across the reviews. Although the maintenance of improvements was found at varying follow-up assessments in one review [34] and at 3-6 months in two additional reviews [26,38], two other reviews reported that evidence was inconsistent with some studies reporting that videoconference interventions may show less longevity in the maintenance of effects up to 6 months compared with face-to-face interventions [21,24]. A final review of mixed modality remote interventions suggested that although inferior to face-to-face formats at shorter follow-up, remote interventions may be more beneficial than face-to-face interventions at longer follow-ups (36 months) [20]. Table 2 presents further details on the clinical outcomes.

XSL•FO
RenderX

Table 2. Clinical effectiveness outcomes.

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
Anxiety					
	Berryhill et al (2019b) [23]	Video-based CBT[a] (n=12); video-based behavioral activation (n=3); video-based ACT[b] (n=1); video-based exposure therapy (n=2); video-based problem-solving therapy (n=1); video-based metacognitive therapy (n=1); multiple modality (n=1)	Face-to-face psychotherapy (n=20); no control (n=1)	In total, 66% (14/21) of studies found statistically significant improvement on validated anxiety measures when videoconferencing psychological therapy was involved. A total of 52% (11/21) of studies reported clinically significant improvements among participants. Controlled study designs comparing face-to-face and videoconferencing psychological therapy sessions (7/10, 70%) found no statistical difference between them.	No combined data were available
	Coughtrey and Pistrang (2018) [26]	Telephone-based CBT (n=2); telephone-based exposure response prevention therapy (n=1); telephone-based behavioral therapy (n=1)	Face-to-face exposure response therapy (n=1); waitlist (n=3)	All 3 RCTs[c] on anxiety reported significant reductions in anxiety symptoms following telephone-delivered intervention (OCD[d]: comparable reductions with face-to-face treatment, maintained over 6-month follow-up. Panic disorder: significant reductions in panic and gains maintained over 3-month follow-up. Transdiagnostic intervention: significant reductions in anxiety sensitivity, panic, social phobia and PTSD[e]). One quasi-experimental study found significant reductions in OCD symptoms as compared with controls maintained at 12-week follow-up.	RCTs: Cohen *d* range 0.34-1.07 (median 0.69; n=2); uncontrolled: Cohen *d*=1.07 (n=1)
	Olthuis et al (2016a) [35]	Internet CBT with therapist email or telephone support (n=37); internet behavioral therapy with exposure (n=1)	Waitlist or attentional control (n=20); face to face (n=7); other internet-based therapy (n=6); multiple control groups (n=5)	Versus control: therapist-supported iCBT[f] showed significantly larger improvements in anxiety (n=12), disorder-specific anxiety symptom severity (n=30), and general anxiety symptom severity (n=19) at posttreatment as compared with waiting list, attentional control, information only or internet-based discussion group only controls. Versus unguided iCBT: therapist-supported iCBT showed no difference in improvements in anxiety at posttreatment (n=1), disorder-specific anxiety symptom severity at posttreatment (n=5), and general anxiety symptom severity (n=2) at posttreatment compared with unguided self-help iCBT. Versus face-to-face: therapist-supported iCBT showed no difference in improvements in anxiety at posttreatment (n=4) and 6- to 12-month follow-up (n=3), disorder-specific anxiety symptom severity at posttreatment (n=7) and 6- to 12-month follow-up (n=6) and general anxiety symptom severity (n=6) at posttreatment and at 6- to 12-month follow-up (n=5) as compared with face-to-face CBT.	Waitlist, attentional control, information only or internet-based discussion group–only controls at posttreatment: SMD[g]=–1.06 (95% CI –1.29 to –0.82), *P*<.001; face-to-face CBT at posttreatment: SMD=0.06 (95% CI –0.25 to 0.37); *P*=.36 (no difference between iCBT and face-to-face sessions)
PTSD					

XSL•FO
RenderX

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
	Turgoose et al (2018, veterans) [38]	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	In total, 18 studies looked at the clinical effectiveness of teletherapy interventions. All of these studies reported that teletherapy was associated with significant reductions in PTSD symptoms, regardless of the type of intervention used, except one study that only measured anger in veterans with PTSD. Of those studies that used follow-up measures, all but one found these changes to be present at 3 or 6 months following treatment. In total, 67% (12/18) of studies compared teletherapy with in-person interventions. In all, 9 studies concluded that teletherapy was as effective as in-person therapy. Two suggested in-person therapy produced significantly greater reductions in PTSD symptoms (though neither were randomized), and 1 study found that teletherapy was more effective than in-person therapy.	No combined data available
	Olthuis et al (2016b) [36]	Video-based CBT (n=3); video-based cognitive processing therapy (n=3); internet CBT with therapist email or telephone support (n=9); video-based prolonged exposure (n=2); telephone mindfulness (n=1); video-based behavioral activation and exposure (n=1)	Face-to-face (n=8); internet-based supportive counseling (n=1); treatment as usual (n=2); telephone (n=1); self-help iCBT (n=1); waiting list (n=6)	Overall, telehealth interventions showed significant improvement in PTSD symptoms postintervention (n=18), at 3- to 6-month follow-up (n=11), and at 7- to 12-month follow-up (n=3); videoconferencing: in total, 9 studies examined videoconferencing interventions for PTSD. Results showed a significant improvement in PTSD symptoms at postintervention. There was no difference in improvements in PTSD symptoms between telehealth and face-to-face interventions at posttreatment (n=7); however, face-to-face interventions showed a significantly greater improvement at 3- to 6-month follow-up (n=5). Internet-based therapy delivered with telephone or email support: in total, 8 studies examined internet-delivered interventions with telephone or email support. Results showed significant improvements in PTSD symptoms at postintervention. Furthermore, telehealth interventions were found to show a significantly greater improvement in PTSD symptoms as compared with waitlist controls (n=6). There were no data comparing these interventions with face-to-face treatments. No follow-up data were available.	Total: within group: pre- and postintervention: g=0.81 (95% CI 0.65 to 0.97), n=18 (favors telehealth); preintervention to 3- to 6-month follow-up: g=0.78 (95% CI 0.59 to 0.97), n=11 (favors telehealth) preintervention to 7- to 12-month follow-up: g=0.75 (95% CI 0.25 to 1.26), n=3 (favors telehealth); Between group: as compared with waitlist control postintervention: g=0.6 (95% CI 0.51 to 0.86), n=6 (favors telehealth) compared with face-to-face treatment for PTSD postintervention: g=–0.05 (95% CI –0.31 to 0.20), n=7 (no difference) compared with face-to-face treatment for PTSD 3- to 6-month follow-up: g=–0.25 (95% CI –0.44 to –0.07), n=5 (favors face-to-face). Videoconferencing: within group: pre- and postintervention: g=0.71 (95% CI 0.47 to 0.96), n=8 (favors telehealth). Between group: compared with waitlist control postintervention: No data compared with face-to-face treatment for PTSD postintervention: g=–0.05 (95% CI –0.31 to 0.20), n=7 (no difference). Internet-delivered interventions with telephone or email support. Within group: pre- and postintervention: g=0.94 (95% CI 0.69 to 1.20), n=8 (favors telehealth). Between group: compared with waitlist control postintervention: g=0.73 (95% CI 0.56 to 0.91), n=5 (favors telehealth) compared with face-to-face treatment for PTSD postintervention (no data)

XSL•FO
RenderX

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
	Bolton and Dorstyn (2015) [24]	iCBT with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no control (n=4)	Therapist-assisted internet programs. Statistically significant reductions in the severity of depression and anxiety symptoms (including PTSD) were associated with therapist-assisted internet programs in five studies, including significant large reductions in fear reactions, suicidal ideation, social functioning, and insomnia. Treatment effects 1-6 months post-telepsychology were mixed, with both deterioration and continued improvement found in psychological functioning. This included an increased risk of alcohol consumption over time but also a decline in PTSD and depression symptoms in participants using internet programs. Videoconferencing: video-based interventions also produced short-term reductions in affective symptoms; however, face-to-face therapy demonstrated slightly higher treatment gains. The longer-term effectiveness of videoconferencing was reported in only two studies which showed nonsignificant effect sizes at follow-up.	No useful synthesis of data
Depression					
	Harerimana et al (2019, older adults) [20]	Telephone-based (n=3); video-based (n=2); web based (n=1)	Waiting list (n=NR[h]) treatment as usual (n=NR)	Telephone: three studies examined a telephone-based intervention. One study found that a home electronic messaging service reduced emergency room and hospital visits. Another study found that older adult veterans given a combined telephone-based psychotherapy and long-term illness management intervention showed significant reductions in depression as compared with usual care. However, a third study found that adding telecoaching to a web intervention did not significantly improve symptoms compared with providing only the web intervention. Videoconferencing: two studies examined Skype-based videoconferencing interventions with inconsistent results. One study found that depression scores improved significantly from baseline but got worse at the 2-month follow-up. Another found that face-to-face and Skype-based intervention were not significantly different at postintervention and shorter follow-ups, but at 36 months, the telehealth intervention showed significantly larger improvements in symptoms. Web-based CBT: one web-based CBT intervention was effective in reducing depression symptoms (P=.04), even with high rates of attrition.	No combined data available

XSL•FO
RenderX

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
	Berryhill et al (2019a) [22]	Video-based CBT (n=12); video-based behavioral activation (n=5); video-based acceptance and behavioral therapy (n=1); video-based exposure (n=3); video-based metacognitive therapy (n=1); video-based problem-solving therapy (n=2); video-based therapy in multiple modalities (n=9)	Face-to-face psychotherapy (n=16); face-to-face or telephone (n=2); no control (n=15)	In total, 67% (22/33) of studies included reported statistically significant reductions in depressive symptoms following videoconference-based psychotherapy. Most controlled studies reported inconsistent results while comparing face-to-face and video-based psychotherapy.	No combined data available
	Coughtrey and Pistrang (2018) [26]	Telephone-based CBT (n=10)	Telephone emotion-focused therapy (n=1); treatment as usual (n=5); no control (n=4)	In total, 83% (5/6) of RCTs on depression reported a significant reduction in depression symptoms following telephone-delivered CBT (n=3) or IPT[i] (n=2). These studies included people with recurrent depression (n=1), HIV (n=1), multiple sclerosis (n=1), and people from rural Latino communities (n=1). Two RCTs reported follow-up: only one of these found the maintenance of reductions in depressive symptoms. One RCT found that depression symptoms were not significantly reduced in veterans. One quasi-experimental study found significant reductions in depression following telephone-delivered CBT, with similar patterns of change found in the comparison group. Three uncontrolled studies reported statistically significant reductions in depression following telephone-delivered CBT, including people with Parkinson disease (n=1), HIV (n=1), and veterans with depression (n=1).	RCTs: Cohen *d* range: 0.25-1.98 (median 0.58), n=5; uncontrolled: Cohen *d* range: 1.13-1.90 (median 1.25), n=2
	Dorstyn et al (2013; minority ethnicity communities) [21]	Telephone CBT (n=2); telephone supportive counseling (n=1); telephone structural ecosystems therapy (n=1); internet-based CBT with weekly individual sessions (n=2); internet telepsychiatry (n=1); internet supportive counseling and personalized email correspondence (n=1)	Face-to-face (n=1); treatment as usual (n=3); minimal support control or waitlist (n=2); no control (n=2)	Telephone- and internet-mediated services were associated with significant improvements in the measurements of depression, anxiety, quality of life and psychosocial functioning. The review also found that two studies demonstrated similar effects on depression ratings (CES-D[j]) in telephone and face-to-face psychotherapy. Three studies reported longer-term effects of telecounseling, with conflicting findings.	No combined data available

XSL•FO
RenderX

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
Carers of people with dementia (for depressive symptoms)	Lins et al (2014) [31]	Telephone counseling (n=9)	Friendly calls (n=3); treatment as usual (n=6)	Telephone counseling without any additional intervention showed significant reductions in depressive symptoms in three studies; however, two additional studies showed no differences between groups. A study of telephone counseling with video sessions showed reductions in depressive symptoms in the intervention group but these did not significantly differ from the control group. One study found that telephone counseling with video sessions and a workbook showed significant reductions in depressive symptoms. Burden, distress, anxiety, quality of life, satisfaction, and social support outcomes were inconsistent. Results show that it is still unclear whether telephone counseling can reduce caregiver burden.	Telephone counseling only: depressive symptoms: n=3, SMD=0.32 (95% CI 0.01 to 0.63), *P*=.04; burden: n=4, SMD=0.45 (95% CI –0.01 to 0.90), *P*=.05
Substance use disorders	Lin et al (2019) [30]	Video or telephone-based psychotherapy (n=10); remote medication management (n=3; patient presents at local clinic with nurse and is connected to a physician at a distant site via videoconference)	Face-to-face psychotherapy (n=7); telephone (n=2); treatment as usual (n=1); no control (n=3)	Tobacco: videoconferencing interventions were not significantly better than in-person (n=1) or telephone (n=2) conditions in terms of abstinence. Alcohol: no significant difference in alcohol use outcomes as compared with usual treatment (n=1), but lower dropout was reported in the telemental health intervention (n=1). Opioid: no significant difference in abstinence between videoconference-based psychotherapy and in-person psychotherapy for methadone patients (n=2), and no difference in time to abstinence (n=1). Notably, none of the included studies described a noninferiority design that specifically assessed whether the intervention was not significantly worse than usual in-person delivered care. Overall, most studies suggested telemental health interventions were an effective alternative, especially when access to treatment is otherwise limited.	No combined data available
Nonspecific					
	Hassan and Sharif (2019, refugee populations) [29]	Not specified videoconferencing treatment intervention (n=2); video-based CBT (n=7); video-based psychoeducation (n=2); video-based relapse prevention (n=1); video-based treatment management (n=1); video-based evaluation of competency to stand trial (n=1)	Face-to-face (n=14)	Five studies compared remote and face-to-face interventions in symptom reduction. Two reviews found greater improvement in the remote intervention, whereas 3 found no significant difference between the intervention and control groups.	No combined data available
	Norwood et al (2018) [34]	Video-based CBT (n=10)	Face-to-face CBT (n=10)	All 10 studies showed that video-based CBT improved symptom severity. Eight studies offered follow-up data, and the postintervention improvement was maintained in all of them. Symptom reduction in video-based CBT was noninferior to face-to-face sessions across all six studies which offered a face-to-face comparison.	No combined data available

XSL•FO
RenderX

Main diagnosis	Study	Intervention (number of studies)	Comparator (number of studies)	Results	Data
	Drago et al (2016) [27]	Videoconferencing (n=24)	Face-to-face (n=23). No comparator (n=1)	In total, 14 RCTs focused on efficacy of remote psychiatric counseling. There was no difference between treatment outcomes in remote and face-to-face settings.	Videoconferencing versus face-to-face therapy: SMD=–0.11 (95% CI –0.41 to 0.18)
	Garcia-Lizana and Munoiz-Mayorga (2010) [28]	Videoconferencing for diagnosis and follow-up (n=3); video-based evaluation of competency to stand trial (n=1); nonspecific video-based CBT (n=5); video-based psychoeducation and counseling (n=1)	Face-to-face (n=10)	Across seven studies, there was no statistically significant difference between telepsychiatry and face-to-face interventions in symptom reduction. Across three studies, there was no statistically significant difference between telepsychiatry in improvements in quality of life.	No combined data available

[a]CBT: cognitive behavioral therapy.

[b]ACT: acceptance and commitment therapy.

[c]RCTs: randomized controlled trials.

[d]OCD: obsessive-compulsive disorder.

[e]PTSD: posttraumatic stress disorder.

[f]iCBT: internet-based cognitive behavioral therapy.

[g]SMD: standardized mean difference.

[h]NR: not reported.

[i]IPT: interpersonal psychotherapy.

[j]CES-D: Center for Epidemiological Studies-Depression Scale.

Implementation Outcomes

Overview

Implementation outcomes were reported in eight reviews [21,24,25,27,28,30,32,38]. Relevant outcomes included assessment comparability (n=2 reviews), fidelity to intervention and competence of therapists (n=1) [38], patient adherence to intervention (n=3) [21,24,28], patient attendance (n=4) [21,25,30,38], safety (n=2) [24,38], and technical difficulties (n=3) [24,25,38].

Assessment Comparability

Limited evidence from one review suggests that videoconferencing can be used to conduct assessment, which is consistent with face-to-face assessment, with a correlation coefficient of 0.73 (95% CI 0.63-0.83) between the conclusions of videoconference and face-to-face assessments [27]. A review of telephone assessments found lack of properly performed studies on telephone assessments [32].

Fidelity and Competence of Therapists

One review [38] found that fidelity and therapist competence using telemental health had been found to be comparable with face-to-face interventions in 3 studies of interventions for PTSD in veterans.

Patient Adherence to Intervention

From 3 reviews [21,24,28] examining patients' adherence to remote interventions the general consensus was that the comprehension of tasks and completion rates are high during both telephone and video-based CBT. However, another review reported mixed findings, with one of the two included relevant studies reporting better adherence in the face-to-face intervention group for patients with PTSD, whereas another study on patients with depression reported equivalent adherence in remote and face-to-face conditions [28].

Patient Attendance

An increase in uptake and access to care following the introduction of telemental health has been reported in the reviews of depression treatment in older adults [20], PTSD treatment in veterans [38], and substance use disorder treatment [30]. Dropouts tended to be comparable with face-to-face interventions [30,38]. However, one review included a study reporting difficulty in reaching ethnic minority patients with depression [21].

Safety

Patient safety while using remote interventions has been reported only in the reviews of populations with PTSD. Two reviews agreed that safety was acceptable, with one reporting that with correct steps taken, safety could usually be managed in remote settings [38], and another study reported that client safety was deemed satisfactory (however, no further details were provided on this result) [24].

Technical Difficulties

A total of 3 reviews reported technical difficulties, none of which were identified as severe barriers to the implementation

XSL•FO
RenderX

of remote technology. A review of older adults with depression found that four studies reported mistrust in technology [25], while the challenges of a more logistical nature, such as low image resolution and connectivity problems, were reported in a review of video-based PTSD interventions for veterans [38]. Another review reported findings from one included study that participants preferred mobile apps to supplement remotely delivered support [24]. Table 3 presents further details on the implementation outcomes.

XSL•FO
RenderX

Table 3. Implementation outcomes.

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
Assessment comparability						
	Drago et al (2016) [27]	Assessment and treatment	Multiple	Videoconferencing (n=24)	Face-to-face (n=23); no comparator (n=1)	Assessment was found to be highly consistent between remote and face-to-face settings; correlation coefficient=0.73 (95% CI 0.63-0.83).
	Muskens et al (2014) [32]	Assessment	Multiple	Telephone diagnostic interviewing (n=16)	Face-to-face diagnostic interviewing (n=16)	There were a few studies that were properly performed to draw conclusions regarding the comparability of telephone and face-to-face interviews for psychiatric morbidity. Telephone interviewing for research purposes in depression and anxiety may however be a valid method.
Fidelity and competence of therapists	Turgoose et al (2018, veterans) [38]	Treatment	PTSD[a]	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	High levels of fidelity and therapist competence (n=3), with no significant differences compared with face-to-face settings.
Patient adherence to intervention						
	Bolton and Dorstyn (2015) [24]	Treatment	PTSD	Internet-based CBT[b] with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no control (n=4)	Qualitative feedback revealed that the comprehension of the therapy materials was high, with participants completing a set of homework tasks (n=5).
	Dorstyn et al (2013, ethnic minorities) [21]	Treatment	Depression	Telephone CBT (n=2); telephone supportive counseling (n=1); telephone structural ecosystems therapy (n=1); internet-based CBT with weekly individual sessions (n=2); internet telepsychiatry (n=1); internet supportive counseling and personalized email correspondence (n=1)	Face-to-face (n=1); treatment as usual (n=3); minimal support control or waitlist (n=2); no control (n=2)	Most studies reported good treatment adherence with rates of completion of 75-97%.

XSL•FO
RenderX

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Garcia-Lizana and Munoiz-Mayorga (2010) [28]	Assessment and treatment	Multiple	Videoconferencing for diagnosis and follow-up (n=3); video-based evaluation of competency to stand trial (n=1); nonspecific video-based CBT (n=5); video-based psychoeducation and counseling (n=1)	Face-to-face (n=10)	Across two studies, mixed results were found for treatment adherence, with one study finding no difference and another reporting better adherence in the face-to-face group.
Patient attendance						
	Dorstyn et al (2013, ethnic minorities) [21]	Treatment	Depression	Telephone CBT (n=2); telephone supportive counseling (n=1); telephone structural ecosystems therapy (n=1); internet-based CBT with weekly individual sessions (n=2); internet telepsychiatry (n=1); internet supportive counseling and personalized email correspondence (n=1)	Face-to-face (n=1); treatment as usual (n=3); minimal support control or waitlist (n=2); no control (n=2)	One study reported difficulty reaching participants by telephone resulting in fewer sessions completed.
	Christensen et al (2019, older adults) [25]	Treatment	Depression or a range of diagnoses including depression	Video consultations for telepsychiatry (n=21)	Face-to-face (11), no control (10)	Video consultations increased access to care and removed barriers such as having to travel (n=4).
	Lin et al (2019) [30]	Treatment	Substance use disorders	Video- or telephone-based psychotherapy (n=10); remote medication management (n=3; patient presents at local clinic with nurse and are connected to a physician at a distant site via videoconference)	Face-to-face psychotherapy (n=7); telephone (n=2); treatment as usual (n=1); no control (n=3)	Most studies reported increased retention in telemental health groups (n=4); however, no difference in in number of sessions attended was sometimes reported (n=2). One alcohol addiction study reported lower dropout in the telemental health group, and more patients in this group were still in treatment at 6 months and one year. Two opioid addiction studies found that videoconference interventions resulted in a better retention of participants up to one year as compared with those receiving in-person care. Another opioid study found >50% retention at 12 weeks but did not have a comparison group. However, another two studies found no difference between videoconference-delivered psychotherapy and in-person psychotherapy in the number of sessions attended.

XSL•FO
RenderX

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Turgoose et al (2018 veterans) [38]	Treatment	PTSD	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	In the majority of cases, there were no differences between teletherapy and in-person treatments on dropout or attendance. There was some indication that teletherapy may help to increase uptake.
Safety						
	Bolton and Dorstyn (2015) [24]	Treatment	PTSD	Internet-based CBT with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no control (n=4)	Client safety was deemed satisfactory.
	Turgoose et al (2018, veterans) [38]	Treatment	PTSD	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	There might be some occasions where veterans have concerns about exposure tasks due to the lack of physical presence of the therapist; however, overall, it was established that these can be used just as effectively remotely. If appropriate steps are taken to manage safety, then episodes of acute suicidality can also be managed.
Technical difficulties						
	Bolton and Dorstyn (2015) [24]	Treatment	PTSD	Internet-based CBT with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no control (n=4)	Minimal technical difficulties were encountered (n=1); participants reported that they would have preferred different forms of media, for example, a mobile app, to supplement support (n=1).
	Christensen et al (2019, older adults) [25]	Treatment	Depression or a range of diagnoses including depression	Video consultations for telepsychiatry (n=21)	Face-to-face (11), no control (10)	Challenges such as mistrust in technology were reported frequently (n=4).

XSL•FO
RenderX

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Turgoose et al (2018, veterans) [38]	Treatment	PTSD	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	Commonly reported technical difficulties were low-image resolution on videoconferencing technology, not being able to connect, and audio delays.

[a]PTSD: posttraumatic stress disorder.

[b]CBT: cognitive behavioral therapy.

Acceptability Outcomes

Overview

Acceptability outcomes were reported in 10 reviews [20,21,24,25,28-31,34,38]. Relevant outcomes included clinician satisfaction (n=5) [20,28,29,31,38], therapeutic alliance (n=6) [24,25,30,31,34,38], patient satisfaction (n=7) [21,25,28-31,38], and convenience (n=3) [25,30,31].

Clinician Satisfaction

Overall, clinicians tend to report a preference for face-to-face interventions for both assessment and treatment [28,29]. However, some reviews have reported that clinicians find video-based therapies to be acceptable [29,38]. One review of remote interventions for the carers of people with dementia found that counselors felt they might need more support via debriefing following remote counseling sessions. They also reported problems when the reactions of carers could not be ascertained via remote technology along with the feelings of helplessness due to the impersonal nature of remote technology [31]. Health care providers using remote interventions in older adults noted the practical benefits of telehealth [20].

Therapeutic Alliance

Overall, therapeutic alliances were reported to be comparable with face-to-face interventions. However, some patient groups were found to feel more comfortable talking to therapists face to face, if possible, such as older female adults [25] or veterans [38]. A meta-analysis was conducted in one review, which found that although standardized mean differences in alliance ratings were not significantly different, the lower limit of the 95% CI fell outside the prespecified limit of noninferiority, indicating that videoconference interventions may be inferior to face-to-face treatment, likely the result of therapist-rated (but not patient-rated) alliance scores being lower in the videoconference groups [34].

Patient Satisfaction

High patient satisfaction was generally reported across seven reviews, and patients tended to find remote interventions as satisfactory as face-to-face alternatives. This was true in substance use disorders [30], depression [21,25,28,29], PTSD [38], older adults [25], ethnic minorities [21], and carers of patient populations with dementia [31]; however, Hassan and Sharif [29] reported a few studies indicating preference for face-to-face interventions. A review of older people noted that initial skepticism between both service users and providers tended to dissipate following positive experiences of videoconferencing; moreover, with appropriate support and access to technology, patients who had not previously used computers reported positive experiences of video calls [25]. Accepting the need for treatment to be in teletherapy form instead of face-to-face therapy was reported as important in a study of veterans with PTSD [38].

Convenience

Patients reported the benefits of added convenience of therapy sessions at home via remote interventions for both depression [25,31] and substance use disorders [30]. Table 4 presents further details on the acceptability outcomes.

XSL•FO
RenderX

Table 4. Acceptability outcomes.

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
Clinician satisfaction						
	Garcia-Lizana and Munoiz-Mayorga (2010) [28]	Assessment and treatment	Multiple	Videoconferencing for diagnosis and follow-up (n=3); video-based evaluation of competency to stand trial (n=1); non-specific video-based CBT[a] (n=5); video-based psychoeducation and counseling (n=1)	Face-to-face (n=10)	The lowest level of satisfaction was found to be in the videoconferencing group in two studies that examined clinician satisfaction.
	Hassan and Sharif (2019; refugee populations) [29]	Assessment and treatment	Multiple	Not specified videoconferencing treatment intervention (n=2); video-based CBT (n=7); video-based psychoeducation (n=2); video-based relapse prevention (n=1); video-based treatment management (n=1); video-based evaluation of competency to stand trial (n=1)	Face-to-face (n=14)	Clinicians tended to report higher satisfaction in the face-to-face interventions; however, most reported good satisfaction with the videoconference group.
	Turgoose et al (2018; veterans) [38]	Treatment	PTSD[b]	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	One study reported that clinicians delivering therapy found teletherapy acceptable, with no difference with in-person therapies.
	Harerimana et al (2019; older adults) [20]	Treatment	Depression	Telephone based (n=6); video based (n=2); web based (n=1)	Waiting list (NR[c]); treatment as usual (NR)	Health care providers have positive perceptions and notice practical benefits associated with the use of telehealth for delivery of community mental health care (n=1). However, nurses of a telepsychiatry consultation generally did not rate it positively (n=1).
	Lins et al (2014) [31]	Support for carers of people with dementia (depressive symptoms)	Carers of people with dementia (for depressive symptoms)	Telephone counseling (n=9, n=2 reporting implementation outcomes)	Friendly calls (n=3); treatment as usual (n=6)	Spatial distance could be a problem because counselors cannot see the reactions of carers (n=1). Counselors also expressed a need for a debriefing with colleagues after counseling sessions.
Therapeutic alliance						
	Bolton and Dorstyn (2015) [24]	Treatment	PTSD	Internet-based CBT with therapist support via telephone calls, introductory face-to-face meetings, or emails (n=6); video-based CBT (n=5)	Face-to-face (n=5); supportive counseling (n=1); waitlist (n=1); no control (n=4)	Good therapeutic alliance reported (n=5)
	Christensen et al (2019, older adults) [25]	Treatment	Depression or a range of diagnoses including depression	Video consultations for telepsychiatry (n=21)	Face-to-face (11), no control (10)	Video sessions were considered better than telephone sessions because of their similarity to face-to-face sessions (n=2); however, in one study, female patients found videoconferencing interventions more impersonal than face-to-face sessions. One clinician reported reduced communication intensity because of less clear facial movements (n=1).

XSL•FO
RenderX

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Lin et al (2019) [30]	Treatment	Substance use disorders	Video- or telephone-based psychotherapy (n=10); remote medication management (n=3; patient presents at local clinic with nurse and are connected to a physician at a distant site via videoconference)	Face-to-face psychotherapy (n=7); telephone (n=2); treatment as usual (n=1); no control (n=3)	Participant and therapist ratings of therapeutic alliance ratings were high in both videoconference and in-person interventions (n=1).
	Lins et al (2014) [31]	Support for carers of people with dementia (depressive symptoms)	Carers of people with dementia (for depressive symptoms)	Telephone counseling (n=2 reporting implementation outcomes)	Friendly calls (n=3); treatment as usual (n=6)	Counselors can feel frustrated and helpless during telephone counseling because it is relatively impersonal (n=1).
	Norwood et al (2018) [34]	Treatment	Multiple	Video-based CBT (n=10)	Face-to-face CBT (n=10)	Six studies used a face-to-face condition as a control group, with four finding that therapeutic alliance was noninferior in the videoconferencing condition as compared with face-to-face conditions. The remaining two reported that alliance was higher in the face-to-face group, though one reported no difference in participant rated alliance, only significantly higher therapist-rated alliance for the face-to-face group. Standardized mean difference in alliance ratings –0.30 (95% CI –0.67 to 0.07), P=.11, n=4. The lower limit of the 95% CI fell outside the prespecified limit of noninferiority (Cohen d=−0.50), indicating that with respect to working alliance, videoconference interventions were inferior to face-to-face treatment.
	Turgoose et al (2018, veterans) [38]	Treatment	PTSD	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	Although most studies found that alliance was equivalent in teletherapy and in-person conditions, some suggested that veterans may feel more comfortable talking to therapists face-to-face. Challenges in detecting body language were reported, but overall clinicians felt that teletherapy did not affect their ability to establish rapport.
Patient satisfaction						

XSL•FO
RenderX

Outcome	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Christensen et al (2019; older adults) [25]	Treatment	Depression or a range of diagnoses including depression	Video consultations for telepsychiatry (n=21)	Face-to-face (11), no control (10)	High levels of patient satisfaction and acceptability were frequently reported, and there were no significant differences between face-to-face and videoconferencing in RCT[d] studies. Patients preferred the reduced waiting time (n=1). Some patients reported initial skepticism as a reason for preference of face-to-face interventions, however this usually dissipated with use of remote technology.
	Dorstyn et al (2013, ethnic minorities) [21]	Treatment	Depression	Telephone CBT (n=2); telephone supportive counseling (n=1); telephone structural ecosystems therapy (n=1); internet-based CBT with weekly individual sessions (n=2); internet telepsychiatry (n=1); internet supportive counseling and personalized email correspondence (n=1)	Face-to-face (n=1); treatment as usual (n=3); minimal support control or waitlist (n=2); no control (n=2)	Consistent patient satisfaction was reported.
	Garcia-Lizana (2010) [28]	Assessment and treatment	Multiple	Videoconferencing for diagnosis and follow-up (n=3); video-based evaluation of competency to stand trial (n=1); nonspecific video-based CBT (n=5); video-based psychoeducation and counseling (n=1)	Face-to-face (n=10)	Patients generally appeared satisfied with the technology utilized and its quality (n=2). High satisfaction was reported in other studies; however, it is unclear if satisfaction was generated by the program or the technology (n=5).
	Hassan and Sharif (2019; refugee populations) [29]	Assessment and treatment	Multiple	Not specified videoconferencing treatment intervention (n=2); video-based CBT (n=7); video-based psychoeducation (n=2); video-based relapse prevention (n=1); video-based treatment management (n=1); video-based evaluation of competency to stand trial (n=1)	Face-to-face (n=14)	Most studies reported a high satisfaction with videoconference interventions (n=3) or no difference in satisfaction as compared with face-to-face groups (n=3); however, one study reported a lower satisfaction as compared with face-to-face sessions.
	Lin et al (2019) [30]	Treatment	Substance use disorders	Video- or telephone-based psychotherapy (n=10); remote medication management (n=3; patient presents at local clinic with nurse and are connected to a physician at a distant site via videoconference)	Face-to-face psychotherapy (n=7); telephone (n=2); treatment as usual (n=1); no control (n=3)	Satisfaction was generally quite high in videoconference interventions and that participants would recommend the intervention to others.
	Lins et al (2014) [31]	Support for carers of people with dementia (depressive symptoms)	Carers of people with dementia (for depressive symptoms)	Telephone counseling (n=9, n=2 reporting implementation outcomes)	Friendly calls (n=3), treatment as usual (n=6)	Reservations expressed about getting advice from an unknown person (n=1). Both studies reported that carers found the information given helpful and were grateful for it. One study found that telephone counseling helped alleviate loneliness in carers.

XSL•FO
RenderX

Out-come	Study	Assessment or treatment	Main diagnosis	Intervention (number of studies)	Comparator (number of studies)	Results
	Turgoose et al (2018; veterans) [38]	Treatment	PTSD	Video-based exposure (n=10); video-based cognitive processing therapy (n=6); video-based CBT (n=5); mixed interventions (n=11); telephone mindfulness (n=1); video-based behavioral activation (n=2); video-based eye movement desensitization and reprocessing (n=1); video-based anger management (n=2); video-based general coping and psychoeducation interventions (n=3)	Face-to-face (n=41)	Patients found teletherapy and face-to-face treatments equally satisfactory: accepting the need for treatments to be in teletherapy form was shown to be important.
Convenience						
	Christensen et al (2019; older adults) [25]	Treatment	Depression or a range of diagnoses including depression	Video consultations for telepsychiatry (n=21)	Face-to-face (11), no control (10)	Patients reported that video consultations were more relaxing, and it was convenient to stay at home (n=3).
	Lin et al (2019) [30]	Treatment	Substance use disorders	Video or telephone-based Psychotherapy (n=10) remote medication management (n=3; patient presents at local clinic with nurse and are connected to a physician at a distant site via videoconference)	Face-to-face psychotherapy (n=7); telephone (n=2); treatment as usual (n=1); no control (n=3)	Participants found the increased convenience important as they would have had difficulty obtaining the intervention without telemental health (n=1).
	Lins et al (2014) [31]	Support for carers of people with dementia (depressive symptoms)	Carers of people with dementia (for depressive symptoms)	Telephone counseling (n=9, n=2 reporting implementation outcomes)	Friendly calls (n=3); treatment as usual (n=6)	Carers found telephone counseling good because it avoided the stress involved in coordinating an appointment (n=1). Needs for 24-hour counselor availability (n=1).

[a]CBT: cognitive behavioral therapy.

[b]PTSD: posttraumatic stress disorder.

[c]NR: not recorded.

[d]RCT: randomized controlled trial.

Cost-Effectiveness

A total of 2 reviews presented conclusions regarding the economic impact of telepsychiatry [29,33]. One review concluded that telepsychiatry can be cost-effective as compared with face-to-face interventions, particularly in rural areas where there were lower numbers of consultations required before telepsychiatry becomes more cost-effective (combatting initial equipment costs) [29]. The second review, in which the focus was on the cost-effectiveness of telepsychiatry, reported that 60% (15/25) of the included studies found telepsychiatry programs to be less expensive than standard in-person care, due to savings such as the travel time and reduced need for patients and their families to take time off work. However, the remaining studies included in the review concluded that telepsychiatry programs were more expensive, particularly because of videoconferencing equipment costs (n=8) or found no difference in costs (n=1). The review also found a large range in reported costs, with, for example, a long-term delivery of telepsychiatry for veterans ranging from US $930 to US $2116 per patient. The cost-effectiveness analyses within the review suggested that telepsychiatry was less cost-effective than face-to-face alternatives. Accordingly, the review concluded that variation was due to a large disparity in the reporting of costs, for example, whether personnel costs or initial equipment costs were included, and that there remains a need for future efforts to determine the cost-effectiveness of different forms of telepsychiatry, particularly for different disorders and applications of remote technology (eg, consultation vs therapy). In addition, a third review [21] examined health service utilization, which impacts cost-effectiveness. They found that the rates of antidepressant and health service utilization were similar in the 3 months following both telephone and web-based counseling as compared with usual care or face-to-face controls.

Guidelines

Only one review [37] of guidelines for telemental health met the inclusion criteria. This review comprehensively summarizes the guidance published to date, including guidance on decisions about the appropriateness of e–mental health, ensuring the competence of mental health professionals, legal and regulatory issues, confidentiality, professional boundaries, and crisis intervention. Recommendations from 19 guidelines were characterized as either firm (50% or more of the guidelines recommended this) or tentative (fewer than 50% of the guidelines recommended this). The review identified the following as firm recommendations: ensuring that the remote interventions were appropriate for the needs of individual

XSL•FO
RenderX

patients and within the boundaries of therapist competence, as well as adhering to laws and regulations; maintaining confidentiality and seeking informed consent (including for specific aspects of remote appointments such as data security); and ensuring that geographically accessible in-person clinical support is available in the case of a crisis or emergency. Guidelines suggested a higher risk of harm for people with cognitive impairments and psychotic disorders but did not provide stronger recommendations on how to adapt delivery to these populations. Furthermore, a minority of guidelines discussed remote technology in young people, with the main message being the importance of checking consent with both the patient and parent. A full summary of the recommendations from the review can be found in Multimedia Appendix 5.

Discussion

Principal Findings

Our umbrella review retrieved a variety of recent relevant systematic reviews on which the future planning of telemental health implementation can be usefully drawn. Across the 19 reviews included in this umbrella review, the results suggest that the remote forms of assessment and intervention can produce at least moderate decreases in symptom severity for people affected by a variety of mental health conditions. Arguments are strongest for videoconferencing interventions, with multiple reviews concluding that outcomes appear comparable with face-to-face interventions at the end of treatment. However, at present, conclusions regarding longer-term results remain uncertain, whereas some reviews have reported the maintenance of positive effects up to six months for both videoconference- and telephone-based interventions, other reviews have suggested that effects are less long-lasting than for face-to-face interventions and the amount of evidence on which to base this assessment is limited. An avenue for future research could be to further examine possible differences between settings in which longer-term benefits have been found compared with those which did not find this.

Reviews also suggest that remote interventions are acceptable to service users participating in studies who tend to report being as satisfied as in face-to-face interventions. This, along with reports that participants adhere to telemental health-based interventions at similar rates to face-to-face interventions, is promising with respect to adaptations during the COVID-19 crisis and for the future, but the reviews tend to relate to small-scale and carefully planned implementations of telemental health with volunteer participants, rather than to large-scale emergency implementations, as in the current crisis. Conclusions drawn from the reviews regarding specific factors that could impact patient satisfaction should also be noted. For example, a report of reduced skepticism following positive experiences [25], if confirmed, suggests that induction sessions and other methods of familiarizing patients with technology and making it accessible to them may improve acceptability and uptake. Clinician satisfaction varied more with reviews tending to conclude that although remote interventions may be acceptable, clinicians usually prefer face-to-face interventions. This may be related to the reports in some reviews that the clinician ratings of therapeutic alliance are poorer with telemental health [34,277]. Despite this, patients tend to feel that the alliance is on par with face-to-face interventions [30,34,38]. There is some suggestion that training and more experience with video and telephone-based technology for intervention delivery may alleviate this concern among therapists [277]; however, staff reports following an increased uptake in the COVID crisis appear to suggest continued concerns about rapport [2]. In the future, a thorough exploration of the exact reasons for acceptability and adherence would benefit the evidence base, for example, certain contextual factors such as appointment type, mood, and time of day may have a substantial influence on patient satisfaction and should be explored further.

Evidence yielded by reviews on the important questions of whether assessments appeared accurate and comprehensive and whether treatment was delivered as intended was limited. A total of 2 reviews examined the comparability of remote versus face-to-face assessment, with one review finding a good correlation between assessments and another finding that there was insufficient high-quality evidence published thus far to draw accurate and meaningful conclusions [27,32]. Regarding fidelity, we found one review that reported good therapist fidelity and competence in remotely delivered interventions in the context of service delivery for veterans with PTSD [38], but systematic reviews do not appear to yield evidence as to whether high fidelity and quality is achieved with telemental health interventions. Standardized training rooted in evidence will be important in ensuring high-quality intervention delivery and overcoming self-doubt among clinicians in delivering remote interventions [37,277,278].

A crucial question regarding the rapid adoption of remote technologies during the pandemic has been whether service users may drop out of or be excluded from care as a result. A minority of the reviews included relevant data, most of which were relatively reassuring. Reviews reported that remote interventions were convenient and that examining uptake reported an increase. When examined, retention was also comparable with face-to-face treatment [30,38]. Reports of technological difficulties were reassuringly few across reviews, although this may be more easily achieved with well-planned, smaller-scale implementations of telemental health that characterize research studies than with larger-scale implementation. However, one aspect of remote delivery on which reviews did not generally report is the risk of complete digital exclusion for patients who may not have the skills or resources to engage with remote therapy or assessments [1,2]. The implementation of telemental health across service systems is only likely to be beneficial if there are clear plans to prevent patients with limited access to technology from being at a disadvantage [279,280], whether by supporting them to engage with remote care or ensuring that equivalent care is available face-to-face.

Digital exclusion may result in the exacerbation of existing inequalities where already disadvantaged groups, such as older adults, people with sensory or cognitive impairment, or members of some BAME groups are at a greater risk of exclusion [1,9,281]. Some reviews have examined this issue [20,25]. A single review by Dorstyn et al [21] reported that members of

XSL•FO
RenderX

predominantly North American ethnic minority communities with depression benefited from telecounseling. A more substantial evidence base is thus urgently required to evaluate the risk of exacerbating ethnic inequalities in mental health care access through telemental health adoption. Furthermore, many have argued that the shift to remote working may exclude older adults [25,281]. On the basis of one review [20] suggesting that videoconferencing interventions can be comparable with face-to-face sessions, and another review [25] finding high levels of patient satisfaction, therapeutic alliance, attendance, and convenience, this review suggests that effective remote intervention delivery may be feasible for older adults. This is encouraging as staying at home and avoiding infection during the pandemic is especially desirable for older adults. No reviews were found regarding other subgroups of potential concern, such as people with sensory or cognitive impairments, children and adolescents and their families, or people with comorbid mental and physical health conditions. We also did not find substantial evidence on settings of particular interest, such as mental health inpatient services (including the use of telemental health in compulsory detention processes) and crisis services.

Limitations

The findings of this umbrella review should be considered along with a number of limitations. First, umbrella reviews aim to present an overview of findings from systematic reviews [282], making conclusions reliant on the quality and accuracy of reporting of included reviews and necessarily resulting in some loss of nuance when findings are pooled. For example, information regarding follow-up periods was not always clear in the conclusions presented in reviews, making a thorough examination of the longevity of benefits difficult. Although we included only reviews considered to be systematic (defined here as searching at least three databases and conducting a quality assessment when synthesizing quantitative data), it was apparent from our quality assessment that the majority of reviews lacked several attributes or characteristics of a high-quality review with robust conclusions, such as prespecified protocols and duplicate study selection. However, our aim was to gain a rapid overview, especially relevant to the current and future rapid implementation of telemental health, of the extent of supporting evidence available in previous literature regarding telemental health: the umbrella review method provides a useful route to achieving this. The inclusion of systematic reviews focused on methods other than randomized controlled trials and on guidance further increases the methodological variability of included reviews and studies, but it is a choice we have made to maximize the retrieval of material from which real-world important lessons can be learned regarding feasibility, acceptability, and implementation barriers and facilitators [283].

This review also aimed to summarize the outcomes related to the cost-effectiveness of remote delivery. We found only two reviews that summarized this outcome and only one that did so comprehensively. It is important that further work should be done to establish the cost-effectiveness of different forms of telemental health, for different patient groups, but there is a significant gap in the literature, despite efficiency being presented as one of the arguments made to support remote interventions [284].

Finally, this review aimed to summarize the literature published before the COVID-19 pandemic to identify evidence relevant to the current context and the recovery from the pandemic. However, the pandemic has given rise to a much more extensive switch to telemental health than previously, meaning that not all conclusions may be generalized to *the new normal*. In particular, the evidence retrieved in this review tends not to relate to the implementation of telemental health across whole catchment areas and does not yield much evidence relevant to currently highly salient issues such as risks of digital exclusion or exacerbation of mental health inequalities and economic disadvantage, which may well be exacerbated as a result of COVID-19 [1,2]. The conclusions of this review should be supplemented with the further scrutiny of the adoption of telemental health within the context of these societal changes, for example through discussion with people with lived experience (Textbox 1).

XSL•FO
RenderX

Textbox 1. Lived experience commentary.

Lived Experience Commentary (by SM From South London Applied Research Collaboration)

As a somewhat long-term user of remote working to access assessment and treatment for a mental disorder, I found it fascinating to read this umbrella review. My experience of remote working has unsurprisingly arisen out of the spatial distancing restrictions imposed upon us all as a result of the SARS-CoV2 outbreak. No surprises there!

My personal experience has been, that although remote working is better than nothing, it isn't in general as good as face-to-face working. Through my patient and public involvement work within my local Trust, I am aware that many fellow patients and carers are also of this view; remote working is okay as an interim measure, but could never replace the therapeutic alliance that can be achieved through sharing space as well as thoughts, feelings, and experiences.

I found it very interesting that the umbrella review found that, in general, clinicians have concerns about remote working as the vast majority of clinicians who I have liaised with via nonlocal advisory groups and PPI have voiced the opinion that they prefer assessing and delivering talking therapy remotely as they find it quicker and more efficient and means they don't have to waste time traveling to different venues. This is understandable; however, for patients who have never met a clinical assessor or therapist before, it can be somewhat alienating to consult with a voice over a phone or an image on a smartphone rather than a present living breathing human being.

One of the main questions which came to mind as I read through the umbrella review was whether it was clarified in any of the individual reviews what service users meant or interpreted remote interventions being satisfactory or acceptable; better than nothing, okay, fine, what? It would have been useful to clarify this.

Another interesting point was that "reports of technological difficulties were reassuringly few across reviews." Maybe it's just me, but I have experienced regular and sometimes quite disastrous "technological difficulties," being locked out of virtual meetings, poor audio and visual reception, loss of signal, not being able to access Wi-Fi, etc. This isn't just annoying but can be incredibly frustrating if you are relying on a therapy session to help you to manage difficult matters in your life, including the mass COVID-19 related isolation and loneliness.

Regarding guidelines for remote working, I am glad that the umbrella review highlighted the gap in guideline provision for remote working with young people. A child psychiatrist I know well has found it very frustrating to have to rely on remote means of assessing young patients and has deep concerns about various risk and safety issues.

Overall, I think the umbrella review raises many pertinent questions and issues, and I hope that at some time in the future, there will be another review of the research literature that will begin in time to proliferate with regard to peoples' experiences of using remote mental health assessments and interventions during this time of COVID-19.

Conclusions

The research across a range of mental health conditions suggests that telemental health is potentially an effective, feasible, and acceptable tool for providing mental health treatment, at least when interventions are relatively well-designed and well-planned, as has been the case in research studies. Comparability in terms of symptom improvement and satisfaction relative to face-to-face methods suggests the move to telemental health to sustain mental health services during the pandemic has been a reasonable one; however, the context of this emergency implementation has been very different from most research studies. Further research should seek to build on existing evidence for establishing the long-term effectiveness and cost-effectiveness of telemental health in a range of groups and settings, such as including children and young people and inpatient acute services and focusing on issues of inclusion and reach. A further question on which further evidence would be highly desirable is the extent to which digital exclusion can be remedied, including the examination of interventions designed to include those with limited previous digital resources or skills. Future planning for telemental health implementation should draw on previous research evidence, often acquired in relatively small-scale studies, and on experiences of trying to engage large service user populations and most of the mental health workforce with remote technology delivery during the COVID-19 pandemic.

Acknowledgments

This paper presents independent research commissioned and funded by the National Institute for Health Research (NIHR) Policy Research Programme conducted by the NIHR Policy Research Unit in Mental Health. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or its arm's length bodies, or other government departments.

This study was supported by the NIHR Applied Research Collaboration (ARC) South London at King's College Hospital National Health Service (NHS) Foundation Trust. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

LG is funded by King's Improvement Science, which is part of the NIHR ARC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. King's Improvement Science is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London, and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity, and Maudsley Charity. JW is funded by the Maudsley Charity. FG is in part supported by the NIHR's Biomedical Research Centre at South London and

XSL•FO
RenderX

Maudsley NHS Foundation Trust and King's College London, the Maudsley Charity, and NIHR ARC South London at King's College Hospital NHS Foundation Trust.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 29 KB - jmir_v23i7e26492_app1.docx]

Multimedia Appendix 2
Study overlap.
[DOCX File , 93 KB - jmir_v23i7e26492_app2.docx]

Multimedia Appendix 3
Included studies in reviews.
[DOCX File , 30 KB - jmir_v23i7e26492_app3.docx]

Multimedia Appendix 4
Detailed quality assessment.
[DOCX File , 27 KB - jmir_v23i7e26492_app4.docx]

Multimedia Appendix 5
Guideline recommendations.
[DOCX File , 22 KB - jmir_v23i7e26492_app5.docx]

References

1. Rains LS, Johnson S, Barnett P, Steare T, Needle J, Carr S, COVID-19 Mental Health Policy Research Unit Group. Early impacts of the COVID-19 pandemic on mental health care and on people with mental health conditions: framework synthesis of international experiences and responses. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):13-24 [FREE Full text] [doi: 10.1007/s00127-020-01924-7] [Medline: 32804258]
2. Johnson S, Dalton-Locke C, Juan NV, Foye U, Oram S, Papamichail A, COVID-19 Mental Health Policy Research Unit Group. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):25-37 [FREE Full text] [doi: 10.1007/s00127-020-01927-4] [Medline: 32857218]
3. Daigle P, Rudnick A. Shifting to remotely delivered mental health care: quality improvement in the COVID-19 pandemic. Psychiatry Int 2020 Sep 22;1(1):31-35. [doi: 10.3390/psychiatryint1010005]
4. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824. [doi: 10.1016/s2215-0366(20)30307-2]
5. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
6. Yellowlees P, Chan SR, Parish MB. The hybrid doctor-patient relationship in the age of technology - telepsychiatry consultations and the use of virtual space. Int Rev Psychiatry 2015 Oct 23;27(6):476-489. [doi: 10.3109/09540261.2015.1082987] [Medline: 26493089]
7. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
8. Aboujaoude E. Telemental health: why the revolution has not arrived. World Psychiatry 2018 Oct 07;17(3):277-278 [FREE Full text] [doi: 10.1002/wps.20551] [Medline: 30192097]
9. Primm A, Vasquez M, Mays R, Sammons-Posey D, McKnight-Eily L, Presley-Cantrell L, et al. The role of public health in addressing racial and ethnic disparities in mental health and mental illness. Prev Chronic Dis 2010 Jan;7(1):A20 [FREE Full text] [Medline: 20040235]
10. Luxton DD, Pruitt LD, Osenbach JE. Best practices for remote psychological assessment via telehealth technologies. Prof Psychol Res Pract 2014 Feb;45(1):27-35. [doi: 10.1037/a0034547]
11. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019 Feb 11;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]

XSL•FO
RenderX

12. Tullio V, Perrone G, Bilotta C, Lanzarone A, Argo A. Psychological support and psychotherapy via digital devices in Covid-19 emergency time: some critical issues. Med Leg J 2020 Jul 03;88(2):73-76. [doi: 10.1177/0025817220926942] [Medline: 32490713]
13. Pieper D, Antoine S, Mathes T, Neugebauer EA, Eikermann M. Systematic review finds overlapping reviews were not mentioned in every other overview. J Clin Epidemiol 2014 Apr;67(4):368-375. [doi: 10.1016/j.jclinepi.2013.11.007] [Medline: 24581293]
14. World Health Organization, Alliance for Health Policy and Systems Research. In: Tricco A, Langlois E, Straus S, editors. Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide. Geneva: World Health Organization; 2017:1-119.
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health 2018 Aug 13;21(3):95-100. [doi: 10.1136/ebmental-2018-300014] [Medline: 30006442]
17. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Br Med J 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
18. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7(1):10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
19. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007 Dec 26;2(12):e1350 [FREE Full text] [doi: 10.1371/journal.pone.0001350] [Medline: 18159233]
20. Harerimana B, Forchuk C, O'Regan T. The use of technology for mental healthcare delivery among older adults with depressive symptoms: a systematic literature review. Int J Ment Health Nurs 2019 Jun 21;28(3):657-670. [doi: 10.1111/inm.12571] [Medline: 30666762]
21. Dorstyn DS, Saniotis A, Sobhanian F. A systematic review of telecounselling and its effectiveness in managing depression amongst minority ethnic communities. J Telemed Telecare 2013 Oct 17;19(6):338-346. [doi: 10.1177/1357633x13501767]
22. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing psychotherapy and depression: a systematic review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
23. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]
24. Bolton A, Dorstyn D. Telepsychology for posttraumatic stress disorder: a systematic review. J Telemed Telecare 2015 Feb 22;21(5):254-267. [doi: 10.1177/1357633x15571996]
25. Christensen LF, Moller AM, Hansen JP, Nielsen CT, Gildberg FA. Patients' and providers' experiences with video consultations used in the treatment of older patients with unipolar depression: a systematic review. J Psychiatr Ment Health Nurs 2020 Jun 26;27(3):258-271. [doi: 10.1111/jpm.12574] [Medline: 31677331]
26. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: a systematic review. J Telemed Telecare 2016 Dec 30;24(2):65-74. [doi: 10.1177/1357633x16686547]
27. Drago A, Winding T, Antypa N. Videoconferencing in psychiatry, a meta-analysis of assessment and treatment. Eur Psychiatry 2016 Aug 23;36:29-37. [doi: 10.1016/j.eurpsy.2016.03.007] [Medline: 27311105]
28. García-Lizana F, Muñoz-Mayorga I. What about telepsychiatry? Prim Care Companion J Clin Psychiatry 2010 Mar 25:A. [doi: 10.4088/pcc.09m00831whi]
29. Hassan A, Sharif K. Efficacy of telepsychiatry in refugee populations: a systematic review of the evidence. Cureus 2019 Jan 30;11(1):e3984 [FREE Full text] [doi: 10.7759/cureus.3984] [Medline: 30972263]
30. Lin L, Casteel D, Shigekawa E, Weyrich MS, Roby DH, McMenamin SB. Telemedicine-delivered treatment interventions for substance use disorders: a systematic review. J Subst Abuse Treat 2019 Jun;101:38-49. [doi: 10.1016/j.jsat.2019.03.007] [Medline: 31006553]
31. Lins S, Rücker G, Motschall E, Langer G, Antes G, Meyer G. Efficacy and experiences of telephone counselling for informal carers of people with dementia. Cochrane Database Syst Rev 2011:CD009126. [doi: 10.1002/14651858.CD009126]
32. Muskens EM, Lucassen P, Groenleer W, van Weel C, Oude Voshaar R, Speckens A. Psychiatric diagnosis by telephone: is it an opportunity? Soc Psychiatry Psychiatr Epidemiol 2014 Oct 15;49(10):1677-1689. [doi: 10.1007/s00127-014-0861-9] [Medline: 24632782]
33. Naslund JA, Mitchell LM, Joshi U, Nagda D, Lu C. Economic evaluation and costs of telepsychiatry programmes: a systematic review. J Telemed Telecare 2020 Aug 03:1357633X2093891. [doi: 10.1177/1357633x20938919]
34. Norwood C, Moghaddam NG, Malins S, Sabin-Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: a systematic review and noninferiority meta-analysis. Clin Psychol Psychother 2018 Nov 16;25(6):797-808. [doi: 10.1002/cpp.2315] [Medline: 30014606]

XSL•FO
RenderX

35. Olthuis J, Watt M, Bailey K, Hayden J, Stewart S. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565 [FREE Full text] [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
36. Olthuis JV, Wozney L, Asmundson GJ, Cramm H, Lingley-Pottie P, McGrath PJ. Distance-delivered interventions for PTSD: a systematic review and meta-analysis. J Anxiety Disord 2016 Dec;44:9-26. [doi: 10.1016/j.janxdis.2016.09.010] [Medline: 27697658]
37. Sansom-Daly UM, Wakefield CE, McGill BC, Wilson HL, Patterson P. Consensus among international ethical guidelines for the provision of videoconferencing-based mental health treatments. JMIR Ment Health 2016 May 18;3(2):e17 [FREE Full text] [doi: 10.2196/mental.5481] [Medline: 27192931]
38. Turgoose D, Ashwick R, Murphy D. Systematic review of lessons learned from delivering tele-therapy to veterans with post-traumatic stress disorder. J Telemed Telecare 2017 Sep 29;24(9):575-585. [doi: 10.1177/1357633x17730443]
39. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A Product From the ESRC Methods Programme Version. 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2021-06-26]
40. Aburizik A, Dindo L, Kaboli P, Charlton M, Dawn K, Turvey C. A pilot randomized controlled trial of a depression and disease management program delivered by phone. J Affect Disord 2013 Nov;151(2):769-774. [doi: 10.1016/j.jad.2013.06.028] [Medline: 23871127]
41. Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada BM, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. Depress Anxiety 2016 May;33(5):415-423. [doi: 10.1002/da.22476] [Medline: 26864655]
42. Acierno R, Knapp R, Tuerk P, Gilmore AK, Lejuez C, Ruggiero K, et al. A non-inferiority trial of Prolonged Exposure for posttraumatic stress disorder: in person versus home-based telehealth. Behav Res Ther 2017 Feb;89:57-65 [FREE Full text] [doi: 10.1016/j.brat.2016.11.009] [Medline: 27894058]
43. Ahmed SN, Mann C, Sinclair DB, Heino A, Iskiw B, Quigley D, et al. Feasibility of epilepsy follow-up care through telemedicine: a pilot study on the patient's perspective. Epilepsia 2008 Apr;49(4):573-585 [FREE Full text] [doi: 10.1111/j.1528-1167.2007.01464.x] [Medline: 18076644]
44. Amarendran V, George A, Gersappe V, Krishnaswamy S, Warren C. The reliability of telepsychiatry for a neuropsychiatric assessment. Telemed J E Health 2011 Apr;17(3):223-225. [doi: 10.1089/tmj.2010.0144] [Medline: 21443440]
45. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461. [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
46. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost LG. Internet-based self-help versus one-session exposure in the treatment of spider phobia: a randomized controlled trial. Cogn Behav Ther 2009;38(2):114-120. [doi: 10.1080/16506070902931326] [Medline: 20183690]
47. Andersson G, Paxling B, Roch-Norlund P, Östman G, Norgren A, Almlöv J, et al. Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom 2012;81(6):344-355 [FREE Full text] [doi: 10.1159/000339371] [Medline: 22964540]
48. Andersson G, Carlbring P, Furmark T, S. O. F. I. E. Research Group. Therapist experience and knowledge acquisition in internet-delivered CBT for social anxiety disorder: a randomized controlled trial. PLoS One 2012;7(5):e37411 [FREE Full text] [doi: 10.1371/journal.pone.0037411] [Medline: 22649526]
49. Andersson G, Waara J, Jonsson U, Malmaeus F, Carlbring P, Ost LG. Internet-based exposure treatment versus one-session exposure treatment of snake phobia: a randomized controlled trial. Cogn Behav Ther 2013;42(4):284-291. [doi: 10.1080/16506073.2013.844202] [Medline: 24245707]
50. Arnaert A, Klooster J, Chow V. Attitudes toward videotelephones: an exploratory study of older adults with depression. J Gerontol Nurs 2007 Sep;33(9):5-13. [doi: 10.3928/00989134-20070901-02] [Medline: 17899995]
51. Aziz MA, Kenford S. Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. J Psychiatr Pract 2004 Sep;10(5):307-313. [doi: 10.1097/00131746-200409000-00004] [Medline: 15361745]
52. Baca CT, Manuel JK. Satisfaction with long-distance motivational interviewing for problem drinking. Addict Disord Their Treat 2007;6(1):39-41 [FREE Full text] [doi: 10.1097/01.adt.0000210708.57327.28]
53. Barrera-Valencia C, Benito-Devia AV, Vélez-Álvarez C, Figueroa-Barrera M, Franco-Idárraga SM. [Cost-effectiveness of synchronous vs. asynchronous telepsychiatry in prison inmates with depression]. Rev Colomb Psiquiatr 2017;46(2):65-73. [doi: 10.1016/j.rcp.2016.04.008] [Medline: 28483175]
54. Berger T, Hohl E, Caspar F. Internet-based treatment for social phobia: a randomized controlled trial. J Clin Psychol 2009 Oct;65(10):1021-1035. [doi: 10.1002/jclp.20603] [Medline: 19437505]
55. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]

XSL•FO
RenderX

56. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
57. Bergström J, Andersson G, Ljótsson B, Rück C, Andréewitch S, Karlsson A, et al. Internet-versus group-administered cognitive behaviour therapy for panic disorder in a psychiatric setting: a randomised trial. BMC Psychiatry 2010 Jul 02;10:54 [FREE Full text] [doi: 10.1186/1471-244X-10-54] [Medline: 20598127]
58. Bishop JE, O'Reilly RL, Maddox K, Hutchinson LJ. Client satisfaction in a feasibility study comparing face-to-face interviews with telepsychiatry. J Telemed Telecare 2002;8(4):217-221. [doi: 10.1258/135763302320272185] [Medline: 12217104]
59. Bouchard S, Paquin B, Payeur R, Allard M, Rivard V, Fournier T, et al. Delivering cognitive-behavior therapy for panic disorder with agoraphobia in videoconference. Telemed J E Health 2004;10(1):13-25. [doi: 10.1089/153056204773644535] [Medline: 15104911]
60. Brøndbo H, Mathiassen B, Martinussen M, Heiervang E, Eriksen M, Kvernmo S. Agreement on web-based diagnoses and severity of mental health problems in norwegian child and adolescent mental health services. Clin Pract Epidemiol Ment Health 2012;8:16-21 [FREE Full text] [doi: 10.2174/1745017901208010016] [Medline: 22582083]
61. Brooks E, Novins DK, Noe T, Bair B, Dailey N, Lowe J, et al. Reaching rural communities with culturally appropriate care: a model for adapting remote monitoring to American Indian veterans with posttraumatic stress disorder. Telemed J E Health 2013 Apr;19(4):272-277. [doi: 10.1089/tmj.2012.0117] [Medline: 23451811]
62. Burke WJ, Roccaforte WH, Wengel SP, Conley DM, Potter JF. The reliability and validity of the Geriatric Depression Rating Scale administered by telephone. J Am Geriatr Soc 1995 Jun;43(6):674-679. [doi: 10.1111/j.1532-5415.1995.tb07205.x] [Medline: 7775729]
63. Butler TN, Yellowlees P. Cost analysis of store-and-forward telepsychiatry as a consultation model for primary care. Telemed J E Health 2012;18(1):74-77. [doi: 10.1089/tmj.2011.0086] [Medline: 22085113]
64. Cacciola JS, Alterman AI, Rutherford MJ, McKay JR, May DJ. Comparability of telephone and In-person structured clinical interview for DSM-III-R (SCID) diagnoses. Assessment 1999 Sep;6(3):235-242. [doi: 10.1177/107319119900600304] [Medline: 10445961]
65. Carlbring P, Nilsson-Ihrfelt E, Waara J, Kollenstam C, Buhrman M, Kaldo V, et al. Treatment of panic disorder: live therapy vs. self-help via the internet. Behav Res Ther 2005 Oct;43(10):1321-1333. [doi: 10.1016/j.brat.2004.10.002] [Medline: 16086983]
66. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/ajp.2006.163.12.2119] [Medline: 17151163]
67. Carlbring P, Gunnarsdóttir M, Hedensjö L, Andersson G, Ekselius L, Furmark T. Treatment of social phobia: randomised trial of internet-delivered cognitive-behavioural therapy with telephone support. Br J Psychiatry 2007 Feb;190:123-128. [doi: 10.1192/bjp.bp.105.020107] [Medline: 17267928]
68. Carlson LE, Lounsberry JJ, Maciejewski O, Wright K, Collacutt V, Taenzer P. Telehealth-delivered group smoking cessation for rural and urban participants: feasibility and cessation rates. Addict Behav 2012 Jan;37(1):108-114. [doi: 10.1016/j.addbeh.2011.09.011] [Medline: 21968227]
69. Cernvall M, Carlbring P, Ljungman L, Ljungman G, von Essen L. Internet-based guided self-help for parents of children on cancer treatment: a randomized controlled trial. Psychooncology 2015 Sep;24(9):1152-1158 [FREE Full text] [doi: 10.1002/pon.3788] [Medline: 25728688]
70. Chang BL. Cognitive-behavioral intervention for homebound caregivers of persons with dementia. Nurs Res 1999;48(3):173-182. [doi: 10.1097/00006199-199905000-00007] [Medline: 10337848]
71. Chang BL, Nitta S, Carter PA, Markham YK. Perceived helpfulness of telephone calls - providing support for caregivers of family members with dementia. J Gerontol Nurs 2004 Sep;30(9):14-21. [doi: 10.3928/0098-9134-20040901-05] [Medline: 15471059]
72. Weintraub E, Greenblatt AD, Chang J, Himelhoch S, Welsh C. Expanding access to buprenorphine treatment in rural areas with the use of telemedicine. Am J Addict 2018 Dec;27(8):612-617. [doi: 10.1111/ajad.12805] [Medline: 30265425]
73. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with alzheimer disease and related dementia. Can J Aging 2009 Dec;28(4):323-336. [doi: 10.1017/S0714980809990158] [Medline: 19925698]
74. Choi I, Zou J, Titov N, Dear BF, Li SD, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
75. Choi NG, Wilson NL, Sirrianni L, Marinucci ML, Hegel MT. Acceptance of home-based telehealth problem-solving therapy for depressed, low-income homebound older adults: qualitative interviews with the participants and aging-service case managers. Gerontologist 2014 Aug;54(4):704-713 [FREE Full text] [doi: 10.1093/geront/gnt083] [Medline: 23929664]

XSL•FO
RenderX

76. Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. Depress Anxiety 2014 Aug;31(8):653-661 [FREE Full text] [doi: 10.1002/da.22242] [Medline: 24501015]
77. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
78. Chong J, Moreno F. Feasibility and acceptability of clinic-based telepsychiatry for low-income Hispanic primary care patients. Telemed J E Health 2012 May;18(4):297-304. [doi: 10.1089/tmj.2011.0126] [Medline: 22424078]
79. Clapp JD, Kemp JJ, Cox KS, Tuerk PW. Patterns of change in response to prolonged exposure: implications for treatment outcome. Depress Anxiety 2016 Sep;33(9):807-815. [doi: 10.1002/da.22534] [Medline: 27321062]
80. Conn DK, Madan R, Lam J, Patterson T, Skirten S. Program evaluation of a telepsychiatry service for older adults connecting a university-affiliated geriatric center to a rural psychogeriatric outreach service in Northwest Ontario, Canada. Int Psychogeriatr 2013 Nov;25(11):1795-1800. [doi: 10.1017/S104161021300118X] [Medline: 23870297]
81. Cowain T. Cognitive-behavioural therapy via videoconferencing to a rural area. Aust N Z J Psychiatry 2001 Feb;35(1):62-64. [doi: 10.1046/j.1440-1614.2001.00853.x] [Medline: 11270458]
82. Crippa JA, de Lima Osório F, Del-Ben CM, Filho AS, da Silva Freitas MC, Loureiro SR. Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. Perspect Psychiatr Care 2008 Oct;44(4):241-247. [doi: 10.1111/j.1744-6163.2008.00183.x] [Medline: 18826462]
83. Crowe T, Jani S, Jani S, Jani N, Jani R. A pilot program in rural telepsychiatry for deaf and hard of hearing populations. Heliyon 2016 Mar;2(3):e00077 [FREE Full text] [doi: 10.1016/j.heliyon.2016.e00077] [Medline: 27441259]
84. Davis LL, Burgio LD, Buckwalter KC, Weaver M. A comparison of in-home and telephone-based skill training interventions with caregivers of persons with Dementia. J Ment Health Aging 2004;10(1):31-44 [FREE Full text]
85. De Las Cuevas C, Artiles J, De La Fuente J, Serrano P. Telepsychiatry in the Canary Islands: user acceptance and satisfaction. J Telemed Telecare 2003;9(4):221-224. [doi: 10.1258/135763303322225553] [Medline: 12952693]
86. De Las Cuevas C, Arredondo MT, Cabrera MF, Sulzenbacher HJ, Meise U. Randomized clinical trial of telepsychiatry through videoconference versus face-to-face conventional psychiatric treatment. Telemed J E Health 2006 Jun;12(3):341-350. [doi: 10.1089/tmj.2006.12.341] [Medline: 16796502]
87. Leo J, Lamb K, LaRowe S, Ana E. A brief behavioral telehealth intervention for veterans with alcohol misuse problems in VA primary care. Drug Alcohol Depend 2014 Jul;140:e45. [doi: 10.1016/j.drugalcdep.2014.02.144]
88. Demiris G, Oliver DP, Wittenberg-Lyles E, Washington K. Use of videophones to deliver a cognitive-behavioural therapy to hospice caregivers. J Telemed Telecare 2011;17(3):142-145 [FREE Full text] [doi: 10.1258/jtt.2010.100503] [Medline: 21303934]
89. Dobkin RD, Menza M, Allen LA, Tiu J, Friedman J, Bienfait KL, et al. Telephone-based cognitive-behavioral therapy for depression in Parkinson disease. J Geriatr Psychiatry Neurol 2011 Dec;24(4):206-214 [FREE Full text] [doi: 10.1177/0891988711422529] [Medline: 22228827]
90. DuHamel KN, Mosher CE, Winkel G, Labay LE, Rini C, Meschian YM, et al. Randomized clinical trial of telephone-administered cognitive-behavioral therapy to reduce post-traumatic stress disorder and distress symptoms after hematopoietic stem-cell transplantation. J Clin Oncol 2010 Aug 10;28(23):3754-3761 [FREE Full text] [doi: 10.1200/JCO.2009.26.8722] [Medline: 20625129]
91. Dunstan DA, Tooth SM. Treatment via videoconferencing: a pilot study of delivery by clinical psychology trainees. Aust J Rural Health 2012 Apr;20(2):88-94. [doi: 10.1111/j.1440-1584.2012.01260.x] [Medline: 22435769]
92. Dwight-Johnson M, Aisenberg E, Golinelli D, Hong S, O'Brien M, Ludman E. Telephone-based cognitive-behavioral therapy for Latino patients living in rural areas: a randomized pilot study. Psychiatr Serv 2011 Aug;62(8):936-942 [FREE Full text] [doi: 10.1176/ps.62.8.pss6208_0936] [Medline: 21807834]
93. Egede LE, Acierno R, Knapp RG, Walker RJ, Payne EH, Frueh BC. Psychotherapy for depression in older veterans via telemedicine: effect on quality of life, satisfaction, treatment credibility, and service delivery perception. J Clin Psychiatry 2016 Dec;77(12):1704-1711. [doi: 10.4088/JCP.16m10951] [Medline: 27835713]
94. Egede LE, Gebregziabher M, Walker RJ, Payne EH, Acierno R, Frueh BC. Trajectory of cost overtime after psychotherapy for depression in older Veterans via telemedicine. J Affect Disord 2017 Jan 01;207:157-162. [doi: 10.1016/j.jad.2016.09.044] [Medline: 27721190]
95. Eibl JK, Gauthier G, Pellegrini D, Daiter J, Varenbut M, Hogenbirk JC, et al. The effectiveness of telemedicine-delivered opioid agonist therapy in a supervised clinical setting. Drug Alcohol Depend 2017 Jul 01;176:133-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.01.048] [Medline: 28535455]
96. Eisdorfer C, Czaja SJ, Loewenstein DA, Rubert MP, Argüelles S, Mitrani VB, et al. The effect of a family therapy and technology-based intervention on caregiver depression. Gerontologist 2003 Aug;43(4):521-531 [FREE Full text] [doi: 10.1093/geront/43.4.521] [Medline: 12937331]
97. Elford R, White H, Bowering R, Ghandi A, Maddiggan B, St John K, et al. A randomized, controlled trial of child psychiatric assessments conducted using videoconferencing. J Telemed Telecare 2000;6(2):73-82. [doi: 10.1258/1357633001935086] [Medline: 10824374]

XSL•FO
RenderX

98. Elford DR, White H, St John K, Maddigan B, Ghandi M, Bowering R. A prospective satisfaction study and cost analysis of a pilot child telepsychiatry service in Newfoundland. J Telemed Telecare 2001;7(2):73-81. [doi: 10.1258/1357633011936192] [Medline: 11331044]
99. Engel CC, Litz B, Magruder KM, Harper E, Gore K, Stein N, et al. Delivery of self training and education for stressful situations (DESTRESS-PC): a randomized trial of nurse assisted online self-management for PTSD in primary care. Gen Hosp Psychiatry 2015;37(4):323-328 [FREE Full text] [doi: 10.1016/j.genhosppsych.2015.04.007] [Medline: 25929985]
100. Evans M, Kessler D, Lewis G, Peters TJ, Sharp D. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? Psychol Med 2004 Jan;34(1):157-162. [doi: 10.1017/s0033291703008055] [Medline: 14971636]
101. Finkel S, Czaja SJ, Schulz R, Martinovich Z, Harris C, Pezzuto D. E-care: a telecommunications technology intervention for family caregivers of dementia patients. Am J Geriatr Psychiatry 2007 May;15(5):443-448. [doi: 10.1097/JGP.0b013e3180437d87] [Medline: 17463195]
102. Fitt S, Rees C. Metacognitive therapy for obsessive compulsive disorder by videoconference: a preliminary study. Behav Change 2013 Jan 09;29(4):213-229. [doi: 10.1017/bec.2012.21]
103. Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007 Aug;22(8):1086-1093 [FREE Full text] [doi: 10.1007/s11606-007-0201-9] [Medline: 17492326]
104. Fortney JC, Pyne JM, Mouden SB, Mittal D, Hudson TJ, Schroeder GW, et al. Practice-based versus telemedicine-based collaborative care for depression in rural federally qualified health centers: a pragmatic randomized comparative effectiveness trial. Am J Psychiatry 2013 Apr;170(4):414-425 [FREE Full text] [doi: 10.1176/appi.ajp.2012.12050696] [Medline: 23429924]
105. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
106. Frank F, Ower N, Zech J, Röhrig J, Gräder N, Berger M, et al. Video conference-based psychotherapeutic follow-up treatment - qualitative case study using a CBASP approach. Psychotherapeut 2017 Mar 30;62(4):355-360. [doi: 10.1007/s00278-017-0181-0] [Medline: 10]
107. Franklin CL, Cuccurullo LA, Walton JL, Arseneau JR, Petersen NJ. Face to face but not in the same place: a pilot study of prolonged exposure therapy. J Trauma Dissociation 2017;18(1):116-130. [doi: 10.1080/15299732.2016.1205704] [Medline: 27348462]
108. Frueh BC, Monnier J, Yim EN, Grubaugh AL, Hamner MB, Knapp RG. A randomized trial of telepsychiatry for post-traumatic stress disorder. J Telemed Telecare 2007;13(3):142-147. [doi: 10.1258/135763307780677604] [Medline: 17519056]
109. Frueh BC, Henderson S, Myrick H. Telehealth service delivery for persons with alcoholism. J Telemed Telecare 2005;11(7):372-375. [doi: 10.1258/135763305774472060] [Medline: 16238840]
110. Frueh BC, Monnier J, Grubaugh AL, Elhai JD, Yim E, Knapp R. Therapist adherence and competence with manualized cognitive-behavioral therapy for PTSD delivered via videoconferencing technology. Behav Modif 2007 Nov;31(6):856-866. [doi: 10.1177/0145445507302125] [Medline: 17932240]
111. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
112. Gant JR, Steffen AM, Lauderdale SA. Comparative outcomes of two distance-based interventions for male caregivers of family members with dementia. Am J Alzheimers Dis Other Demen 2007;22(2):120-128 [FREE Full text] [doi: 10.1177/1533317506298880] [Medline: 17545139]
113. Garzón-Maldonado FJ, Gutiérrez-Bedmar M, Serrano-Castro V, Requena-Toro MV, Padilla-Romero L, García-Casares N. An assessesment of telephone assistance systems for caregivers of patients with Alzheimer's disease. Neurologia 2017;32(9):595-601 [FREE Full text] [doi: 10.1016/j.nrl.2016.03.009] [Medline: 27293022]
114. Gerlach-Reinholz W, Drop L, Basic E, Rauchhaus M, Fritze J. [Telephone coaching for depression]. Nervenarzt 2017 Jul;88(7):811-818. [doi: 10.1007/s00115-017-0316-0] [Medline: 28429078]
115. Germain V, Marchand A, Bouchard S, Drouin MS, Guay S. Effectiveness of cognitive behavioural therapy administered by videoconference for posttraumatic stress disorder. Cogn Behav Ther 2009;38(1):42-53. [doi: 10.1080/16506070802473494] [Medline: 19235601]
116. Glueckauf RL, Davis WS, Willis F, Sharma D, Gustafson DJ, Hayes J, et al. Telephone-based, cognitive-behavioral therapy for African American dementia caregivers with depression: initial findings. Rehabil Psychol 2012 May;57(2):124-139. [doi: 10.1037/a0028688] [Medline: 22686551]
117. Godleski L, Cervone D, Vogel D, Rooney M. Home telemental health implementation and outcomes using electronic messaging. J Telemed Telecare 2012 Jan;18(1):17-19. [doi: 10.1258/jtt.2011.100919] [Medline: 22052966]
118. Gonzalez GE, Brossart DF. Telehealth videoconferencing psychotherapy in rural primary care. J Rur Ment Health 2015 Jul;39(3-4):137-152. [doi: 10.1037/rmh0000037]

XSL•FO
RenderX

119. Greene CJ, Morland LA, Macdonald A, Frueh BC, Grubbs KM, Rosen CS. How does tele-mental health affect group therapy process? Secondary analysis of a noninferiority trial. J Consult Clin Psychol 2010 Oct;78(5):746-750. [doi: 10.1037/a0020158] [Medline: 20873910]
120. Greenwood J, Chamberlain C, Parker G. Evaluation of a rural telepsychiatry service. Aust Psychiatry 2004 Sep 01;12(3):268-272. [doi: 10.1080/j.1039-8562.2004.02097.x] [Medline: 15715789]
121. Griffiths L, Blignault I, Yellowlees P. Telemedicine as a means of delivering cognitive-behavioural therapy to rural and remote mental health clients. J Telemed Telecare 2006;12(3):136-140. [doi: 10.1258/135763306776738567] [Medline: 16638234]
122. Gros DF, Yoder M, Tuerk PW, Lozano BE, Acierno R. Exposure therapy for PTSD delivered to veterans via telehealth: predictors of treatment completion and outcome and comparison to treatment delivered in person. Behav Ther 2011 Jun;42(2):276-283. [doi: 10.1016/j.beth.2010.07.005] [Medline: 21496512]
123. Gros DF, Price M, Strachan M, Yuen EK, Milanak ME, Acierno R. Behavioral activation and therapeutic exposure: an investigation of relative symptom changes in PTSD and depression during the course of integrated behavioral activation, situational exposure, and imaginal exposure techniques. Behav Modif 2012 Jul;36(4):580-599 [FREE Full text] [doi: 10.1177/0145445512448097] [Medline: 22679240]
124. Gros DF, Lancaster CL, López CM, Acierno R. Treatment satisfaction of home-based telehealth versus in-person delivery of prolonged exposure for combat-related PTSD in veterans. J Telemed Telecare 2018 Jan;24(1):51-55. [doi: 10.1177/1357633X16671096] [Medline: 27672059]
125. Grubbs KM, Fortney JC, Pyne JM, Hudson T, Moore WM, Custer P, et al. Predictors of initiation and engagement of cognitive processing therapy among veterans with PSTD enrolled in collaborative care. J Trauma Stress 2015 Dec;28(6):580-584. [doi: 10.1002/jts.22049] [Medline: 26625355]
126. Grubbs KM, Fortney JC, Kimbrell T, Pyne JM, Hudson T, Robinson D, et al. Usual care for rural veterans with posttraumatic stress disorder. J Rural Health 2017 Jun;33(3):290-296. [doi: 10.1111/jrh.12230] [Medline: 28112433]
127. Hajebi A, Motevalian A, Amin-Esmaeili M, Hefazi MS, Radgoodarzi R, Rahimi-Movaghar A, et al. Telephone versus face-to-face administration of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, for diagnosis of psychotic disorders. Compr Psychiatry 2012 Jul;53(5):579-583. [doi: 10.1016/j.comppsych.2011.06.001] [Medline: 21820651]
128. Hassija C, Gray MJ. The effectiveness and feasibility of videoconferencing technology to provide evidence-based treatment to rural domestic violence and sexual assault populations. Telemed J E Health 2011 May;17(4):309-315. [doi: 10.1089/tmj.2010.0147] [Medline: 21457012]
129. Hedman E, Andersson G, Ljótsson B, Andersson E, Rück C, Mörtberg E, et al. Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: a randomized controlled non-inferiority trial. PLoS One 2011 Mar 25;6(3):e18001 [FREE Full text] [doi: 10.1371/journal.pone.0018001] [Medline: 21483704]
130. Hernandez-Tejada MA, Zoller JS, Ruggiero KJ, Kazley AS, Acierno R. Early treatment withdrawal from evidence-based psychotherapy for PTSD: telemedicine and in-person parameters. Int J Psychiatry Med 2014;48(1):33-55. [doi: 10.2190/PM.48.1.d] [Medline: 25354925]
131. Hilty DM, Marks S, Wegelin J, Callahan EJ, Nesbitt TS. A randomized, controlled trial of disease management modules, including telepsychiatric care, for depression in rural primary care. Psychiatry (Edgmont) 2007 Feb;4(2):58-65 [FREE Full text] [Medline: 20805900]
132. Himelhoch S, Mohr D, Maxfield J, Clayton S, Weber E, Medoff D, et al. Feasibility of telephone-based cognitive behavioral therapy targeting major depression among urban dwelling African-American people with co-occurring HIV. Psychol Health Med 2011 Mar;16(2):156-165 [FREE Full text] [doi: 10.1080/13548506.2010.534641] [Medline: 21328144]
133. Himle JA, Fischer DJ, Muroff JR, Van Etten ML, Lokers LM, Abelson JL, et al. Videoconferencing-based cognitive-behavioral therapy for obsessive-compulsive disorder. Behav Res Ther 2006 Dec;44(12):1821-1829. [doi: 10.1016/j.brat.2005.12.010] [Medline: 16466688]
134. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
135. Ivarsson D, Blom M, Hesser H, Carlbring P, Enderby P, Nordberg R, et al. Guided internet-delivered cognitive behavior therapy for post-traumatic stress disorder: a randomized controlled trial. Internet Int 2014 Mar;1(1):33-40. [doi: 10.1016/j.invent.2014.03.002]
136. Jaconis M, Santa Ana EJ, Killeen TK, Badour CL, Back SE. Concurrent treatment of PTSD and alcohol use disorder via telehealth in a female Iraq veteran. Am J Addict 2017 Mar;26(2):112-114 [FREE Full text] [doi: 10.1111/ajad.12481] [Medline: 28118514]
137. Jang Y, Chiriboga DA, Molinari V, Roh S, Park Y, Kwon S, et al. Telecounseling for the linguistically isolated: a pilot study with older Korean immigrants. Gerontologist 2014 Apr;54(2):290-296 [FREE Full text] [doi: 10.1093/geront/gns196] [Medline: 23355448]
138. Johnston L, Titov N, Andrews G, Spence J, Dear BF. A RCT of a transdiagnostic internet-delivered treatment for three anxiety disorders: examination of support roles and disorder-specific outcomes. PLoS One 2011;6(11):e28079 [FREE Full text] [doi: 10.1371/journal.pone.0028079] [Medline: 22132216]

XSL•FO
RenderX

139. Jones MD, Etherage JR, Harmon SC, Okiishi JC. Acceptability and cost-effectiveness of military telehealth mental health screening. Psychol Serv 2012 May;9(2):132-143. [doi: 10.1037/a0026709] [Medline: 22662728]
140. Jones DJ, Forehand R, Cuellar J, Parent J, Honeycutt A, Khavjou O, et al. Technology-enhanced program for child disruptive behavior disorders: development and pilot randomized control trial. J Clin Child Adolesc Psychol 2014;43(1):88-101 [FREE Full text] [doi: 10.1080/15374416.2013.822308] [Medline: 23924046]
141. Jong M. Managing suicides via videoconferencing in a remote northern community in Canada. Int J Circumpolar Health 2004 Dec;63(4):422-428. [doi: 10.3402/ijch.v63i4.17759] [Medline: 15709317]
142. Kennedy C, Yellowlees P. A community-based approach to evaluation of health outcomes and costs for telepsychiatry in a rural population: preliminary results. J Telemed Telecare 2000;6 Suppl 1:155-157. [doi: 10.1258/1357633001934492] [Medline: 10794004]
143. Kim SS, Sitthisongkram S, Bernstein K, Fang H, Choi WS, Ziedonis D. A randomized controlled trial of a videoconferencing smoking cessation intervention for Korean American women: preliminary findings. Int J Womens Health 2016;8:453-462 [FREE Full text] [doi: 10.2147/IJWH.S109819] [Medline: 27660494]
144. King VL, Stoller KB, Kidorf MM, Kindbom K, Hursh S, Brady T, et al. Assessing the effectiveness of an Internet-based videoconferencing platform for delivering intensified substance abuse counseling. J Subst Abuse Treat 2009 Apr;36(3):331-338. [doi: 10.1016/j.jsat.2008.06.011] [Medline: 18775625]
145. King VL, Brooner RK, Peirce JM, Kolodner K, Kidorf MS. A randomized trial of web-based videoconferencing for substance abuse counseling. J Subst Abuse Treat 2014 Jan;46(1):36-42 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.009] [Medline: 24035556]
146. Kiropoulos LA, Klein B, Austin DW, Gilson K, Pier C, Mitchell J, et al. Is internet-based CBT for panic disorder and agoraphobia as effective as face-to-face CBT? J Anxiety Disord 2008 Dec;22(8):1273-1284. [doi: 10.1016/j.janxdis.2008.01.008] [Medline: 18289829]
147. Klee A, Stacy M, Rosenheck R, Harkness L, Tsai J. Interest in technology-based therapies hampered by access: a survey of veterans with serious mental illnesses. Psychiatr Rehabil J 2016 Jun;39(2):173-179. [doi: 10.1037/prj0000180] [Medline: 26985680]
148. Klein B, Mitchell J, Abbott J, Shandley K, Austin D, Gilson K, et al. A therapist-assisted cognitive behavior therapy internet intervention for posttraumatic stress disorder: pre-, post- and 3-month follow-up results from an open trial. J Anxiety Disord 2010 Aug;24(6):635-644. [doi: 10.1016/j.janxdis.2010.04.005] [Medline: 20447802]
149. Knaevelsrud C, Brand J, Lange A, Ruwaard J, Wagner B. Web-based psychotherapy for posttraumatic stress disorder in war-traumatized Arab patients: randomized controlled trial. J Med Internet Res 2015 Mar 20;17(3):e71 [FREE Full text] [doi: 10.2196/jmir.3582] [Medline: 25799024]
150. Kobak KA, Williams JB, Jeglic E, Salvucci D, Sharp IR. Face-to-face versus remote administration of the Montgomery-Asberg Depression Rating Scale using videoconference and telephone. Depress Anxiety 2008;25(11):913-919. [doi: 10.1002/da.20392] [Medline: 17941100]
151. Kobak KA, Greist R, Jacobi DM, Levy-Mack H, Greist JH. Computer-assisted cognitive behavior therapy for obsessive-compulsive disorder: a randomized trial on the impact of lay vs. professional coaching. Ann Gen Psychiatry 2015;14:10 [FREE Full text] [doi: 10.1186/s12991-015-0048-0] [Medline: 25722737]
152. Kok RN, van Straten A, Beekman AT, Cuijpers P. Short-term effectiveness of web-based guided self-help for phobic outpatients: randomized controlled trial. J Med Internet Res 2014 Sep 29;16(9):e226 [FREE Full text] [doi: 10.2196/jmir.3429] [Medline: 25266929]
153. Lazzari C, Egan SJ, Rees CS. Behavioral activation treatment for depression in older adults delivered via videoconferencing: a pilot study. Cognit Behav Pract 2011 Nov;18(4):555-565. [doi: 10.1016/j.cbpra.2010.11.009]
154. Lewis C, Roberts N, Vick T, Bisson JI. Development of a guided self-help (GSH) program for the treatment of mild-to-moderate posttraumatic stress disorder (PTSD). Depress Anxiety 2013 Nov;30(11):1121-1128. [doi: 10.1002/da.22128] [Medline: 23670782]
155. Lichstein KL, Scogin F, Thomas SJ, DiNapoli EA, Dillon HR, McFadden A. Telehealth cognitive behavior therapy for co-occurring insomnia and depression symptoms in older adults. J Clin Psychol 2013 Oct;69(10):1056-1065 [FREE Full text] [doi: 10.1002/jclp.22030] [Medline: 24014056]
156. Lightstone AJ, Bailey SK, Voros P. Collaborative music therapy via remote video technology to reduce a veteran's symptoms of severe, chronic PTSD. Arts Health 2015 Apr 19;7(2):123-136. [doi: 10.1080/17533015.2015.1019895]
157. Lindsay JA, Kauth MR, Hudson S, Martin LA, Ramsey DJ, Daily L, et al. Implementation of video telehealth to improve access to evidence-based psychotherapy for posttraumatic stress disorder. Telemed J E Health 2015 Jun;21(6):467-472 [FREE Full text] [doi: 10.1089/tmj.2014.0114] [Medline: 25714664]
158. Littleton H, Buck K, Rosman L, Grills-Taquechel A. From survivor to thriver: a pilot study of an online program for rape victims. Cogn Behav Pract 2012 May 01;19(2):315-327 [FREE Full text] [doi: 10.1016/j.cbpra.2011.04.002] [Medline: 24009410]
159. Litz BT, Engel CC, Bryant RA, Papa A. A randomized, controlled proof-of-concept trial of an Internet-based, therapist-assisted self-management treatment for posttraumatic stress disorder. Am J Psychiatry 2007 Nov;164(11):1676-1683. [doi: 10.1176/appi.ajp.2007.06122057] [Medline: 17974932]

XSL•FO
RenderX

160. Lovell K, Fullalove L, Garvey R, Brooker C. Telephone treatment of obsessive-compulsive disorder. Behav Cogn Psychother 2000 Jan 01;28(1):87-91. [doi: 10.1017/S1352465800000084]
161. Luxton DD, Pruitt LD, O'Brien K, Kramer G. An evaluation of the feasibility and safety of a home-based telemental health treatment for posttraumatic stress in the U.S. military. Telemed J E Health 2015 Nov;21(11):880-886. [doi: 10.1089/tmj.2014.0235] [Medline: 26103565]
162. Luxton DD, Pruitt LD, Wagner A, Smolenski DJ, Jenkins-Guarnieri MA, Gahm G. Home-based telebehavioral health for U.S. military personnel and veterans with depression: a randomized controlled trial. J Consult Clin Psychol 2016 Nov;84(11):923-934. [doi: 10.1037/ccp0000135] [Medline: 27599225]
163. Lyneham HJ, Rapee RM. Agreement between telephone and in-person delivery of a structured interview for anxiety disorders in children. J Am Acad Child Adolesc Psychiatry 2005 Mar;44(3):274-282. [doi: 10.1097/00004583-200503000-00012] [Medline: 15725972]
164. Maieritsch KP, Smith TL, Hessinger JD, Ahearn EP, Eickhoff JC, Zhao Q. Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. J Telemed Telecare 2016 Jun;22(4):238-243. [doi: 10.1177/1357633X15596109] [Medline: 26231819]
165. Malhotra S, Chakrabarti S, Shah R, Gupta A, Mehta A, Nithya B, et al. Development of a novel diagnostic system for a telepsychiatric application: a pilot validation study. BMC Res Notes 2014 Aug 09;7:508 [FREE Full text] [doi: 10.1186/1756-0500-7-508] [Medline: 25106438]
166. Manchanda M, McLaren P. Cognitive behaviour therapy via interactive video. J Telemed Telecare 1998;4 Suppl 1:53-55. [doi: 10.1258/1357633981931452] [Medline: 9640736]
167. Manguno-Mire GM, Thompson JW, Shore JH, Croy CD, Artecona JF, Pickering JW. The use of telemedicine to evaluate competency to stand trial: a preliminary randomized controlled study. J Am Acad Psychiatry Law 2007;35(4):481-489. [Medline: 18086740]
168. Marchand A, Beaulieu-Prévost D, Guay S, Bouchard SK, Drouin M, Germain V. Relative efficacy of cognitive-behavioral therapy administered by videoconference for posttraumatic stress disorder: a six-month follow-up. J Aggress Maltreat Trauma 2011 Apr 14;20(3):304-321 [FREE Full text] [doi: 10.1080/10926771.2011.562479] [Medline: 23687441]
169. Matsuura S, Hosaka T, Yukiyama T, Ogushi Y, Okada Y, Haruki Y, et al. Application of telepsychiatry: a preliminary study. Psychiatry Clin Neurosci 2000 Feb;54(1):55-58 [FREE Full text] [doi: 10.1046/j.1440-1819.2000.00637.x] [Medline: 15558880]
170. McLellan LF, Andrijic V, Davies S, Lyneham HJ, Rapee RM. Delivery of a therapist-facilitated telecare anxiety program to children in rural communities: a pilot study. Behav Change 2017 Aug 22;34(3):156-167. [doi: 10.1017/bec.2017.11]
171. Menon AS, Kondapavalru P, Krishna P, Chrismer JB, Raskin A, Hebel JR, et al. Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill veterans. J Nerv Ment Dis 2001 Jun;189(6):399-401. [doi: 10.1097/00005053-200106000-00009] [Medline: 11434642]
172. Miller L, Weissman M. Interpersonal psychotherapy delivered over the telephone to recurrent depressives. A pilot study. Depress Anxiety 2002;16(3):114-117. [doi: 10.1002/da.10047] [Medline: 12415535]
173. Miller CJ, McInnes DK, Stolzmann K, Bauer MS. Interest in use of technology for healthcare among veterans receiving treatment for mental health. Telemed J E Health 2016 Oct;22(10):847-854. [doi: 10.1089/tmj.2015.0190] [Medline: 26982279]
174. Mitchell JE, Crosby RD, Wonderlich SA, Crow S, Lancaster K, Simonich H, et al. A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2008 May;46(5):581-592 [FREE Full text] [doi: 10.1016/j.brat.2008.02.004] [Medline: 18374304]
175. Modai I, Jabarin M, Kurs R, Barak P, Hanan I, Kitain L. Cost effectiveness, safety, and satisfaction with video telepsychiatry versus face-to-face care in ambulatory settings. Telemed J E Health 2006 Oct;12(5):515-520. [doi: 10.1089/tmj.2006.12.515] [Medline: 17042703]
176. Mohr DC, Likosky W, Bertagnolli A, Goodkin DE, Van Der Wende J, Dwyer P, et al. Telephone-administered cognitive-behavioral therapy for the treatment of depressive symptoms in multiple sclerosis. J Consult Clin Psychol 2000 Apr;68(2):356-361. [doi: 10.1037//0022-006x.68.2.356] [Medline: 10780138]
177. Mohr DC, Hart SL, Julian L, Catledge C, Honos-Webb L, Vella L, et al. Telephone-administered psychotherapy for depression. Arch Gen Psychiatry 2005 Sep;62(9):1007-1014. [doi: 10.1001/archpsyc.62.9.1007] [Medline: 16143732]
178. Mohr DC, Hart SL, Marmar C. Telephone administered cognitive-behavioral therapy for the treatment of depression in a rural primary care clinic. Cogn Ther Res 2006 May 6;30(1):29-37. [doi: 10.1007/s10608-006-9006-0]
179. Mohr DC, Carmody T, Erickson L, Jin L, Leader J. Telephone-administered cognitive behavioral therapy for veterans served by community-based outpatient clinics. J Consult Clin Psychol 2011 Apr;79(2):261-265. [doi: 10.1037/a0022395] [Medline: 21299274]
180. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]

XSL•FO
RenderX

181. Moreno FA, Chong J, Dumbauld J, Humke M, Byreddy S. Use of standard Webcam and Internet equipment for telepsychiatry treatment of depression among underserved Hispanics. Psychiatr Serv 2012 Dec;63(12):1213-1217. [doi: 10.1176/appi.ps.201100274] [Medline: 23026854]
182. Morland LA, Pierce K, Wong MY. Telemedicine and coping skills groups for Pacific Island veterans with post-traumatic stress disorder: a pilot study. J Telemed Telecare 2004;10(5):286-289. [doi: 10.1258/1357633042026387] [Medline: 15494087]
183. Morland A, Greene CJ, Rosen CS, Foy D, Reilly P, Shore J, et al. Telemedicine for anger management therapy in a rural population of combat veterans with posttraumatic stress disorder: a randomized noninferiority trial. J Clin Psychiatry 2010 Jul;71(7):855-863. [doi: 10.4088/JCP.09m05604blu] [Medline: 20122374]
184. Morland LA, Hynes AK, Mackintosh MA, Resick PA, Chard KM. Group cognitive processing therapy delivered to veterans via telehealth: a pilot cohort. J Trauma Stress 2011 Aug;24(4):465-469. [doi: 10.1002/jts.20661] [Medline: 21793047]
185. Morland LA, Raab M, Mackintosh MA, Rosen CS, Dismuke CE, Greene CJ, et al. Telemedicine: a cost-reducing means of delivering psychotherapy to rural combat veterans with PTSD. Telemed J E Health 2013 Oct;19(10):754-759 [FREE Full text] [doi: 10.1089/tmj.2012.0298] [Medline: 23931729]
186. Morland LA, Mackintosh MA, Greene CJ, Rosen CS, Chard KM, Resick P, et al. Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health: a randomized noninferiority clinical trial. J Clin Psychiatry 2014 May;75(5):470-476. [doi: 10.4088/JCP.13m08842] [Medline: 24922484]
187. Morland LA, Mackintosh MA, Rosen CS, Willis E, Resick P, Chard K, et al. Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: a randomized noninferiority tria. Depress Anxiety 2015 Nov;32(11):811-820. [doi: 10.1002/da.22397] [Medline: 26243685]
188. Cullum CM, Hynan LS, Grosch M, Parikh M, Weiner MF. Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. J Int Neuropsychol Soc 2014 Nov;20(10):1028-1033 [FREE Full text] [doi: 10.1017/S1355617714000873] [Medline: 25343269]
189. Nelson EL, Barnard M, Cain S. Treating childhood depression over videoconferencing. Telemed J E Health 2003;9(1):49-55. [doi: 10.1089/153056203763317648] [Medline: 12699607]
190. Neufeld J, Case R. Walk-in telemental health clinics improve access and efficiency: a 2-year follow-up analysis. Telemed J E Health 2013 Dec;19(12):938-941 [FREE Full text] [doi: 10.1089/tmj.2013.0076] [Medline: 24050609]
191. Newby JM, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013 Dec;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
192. Nieminen K, Berg I, Frankenstein K, Viita L, Larsson K, Persson U, et al. Internet-provided cognitive behaviour therapy of posttraumatic stress symptoms following childbirth-a randomized controlled trial. Cogn Behav Ther 2016 Jun;45(4):287-306. [doi: 10.1080/16506073.2016.1169626] [Medline: 27152849]
193. Niles BL, Klunk-Gillis J, Ryngala DJ, Silberbogen AK, Paysnick A, Wolf EJ. Comparing mindfulness and psychoeducation treatments for combat-related PTSD using a telehealth approach. Psychol Trauma Theory Res Pract Policy 2012;4(5):538-547. [doi: 10.1037/a0026161]
194. Nordgren LB, Hedman E, Etienne J, Bodin J, Kadowaki A, Eriksson S, et al. Effectiveness and cost-effectiveness of individually tailored Internet-delivered cognitive behavior therapy for anxiety disorders in a primary care population: a randomized controlled trial. Behav Res Ther 2014 Aug;59:1-11 [FREE Full text] [doi: 10.1016/j.brat.2014.05.007] [Medline: 24933451]
195. O'Reilly R, Bishop J, Maddox K, Hutchinson L, Fisman M, Takhar J. Is telepsychiatry equivalent to face-to-face psychiatry? Results from a randomized controlled equivalence trial. Psychiatr Serv 2007 Jun;58(6):836-843. [doi: 10.1176/ps.2007.58.6.836] [Medline: 17535945]
196. Ojserkis R, Morris B, McKay D. Pediatric obsessive-compulsive disorder. Clinical Case Studies 2013 Sep 25;13(1):68-79. [doi: 10.1177/1534650113504489]
197. Olthuis JV, Watt MC, Mackinnon SP, Stewart SH. Telephone-delivered cognitive behavioral therapy for high anxiety sensitivity: a randomized controlled trial. J Consult Clin Psychol 2014 Dec;82(6):1005-1022. [doi: 10.1037/a0037027] [Medline: 24911423]
198. Paing WW, Weller RA, Dixon TA, Weller EB. Face-to-face versus telephone administration of the parent's version of the children's interview for psychiatric syndromes (P-ChIPS). Curr Psychiatry Rep 2010 Apr;12(2):122-126. [doi: 10.1007/s11920-010-0101-2] [Medline: 20425297]
199. Painter JT, Fortney JC, Austen MA, Pyne JM. Cost-effectiveness of telemedicine-based collaborative care for posttraumatic stress disorder. Psychiatr Serv 2017 Nov 01;68(11):1157-1163. [doi: 10.1176/appi.ps.201600485] [Medline: 28669290]
200. Paulsen AS, Crowe RR, Noyes R, Pfohl B. Reliability of the telephone interview in diagnosing anxiety disorders. Arch Gen Psychiatry 1988 Jan;45(1):62-63. [doi: 10.1001/archpsyc.1988.01800250076009] [Medline: 3337610]
201. Paxling B, Almlöv J, Dahlin M, Carlbring P, Breitholtz E, Eriksson T, et al. Guided internet-delivered cognitive behavior therapy for generalized anxiety disorder: a randomized controlled trial. Cogn Behav Ther 2011;40(3):159-173. [doi: 10.1080/16506073.2011.576699] [Medline: 21770848]

XSL•FO
RenderX

202. Poon P, Hui E, Dai D, Kwok T, Woo J. Cognitive intervention for community-dwelling older persons with memory problems: telemedicine versus face-to-face treatment. Int J Geriatr Psychiatry 2005 Mar;20(3):285-286. [doi: 10.1002/gps.1282] [Medline: 15717335]
203. Price M, Gros DF. Examination of prior experience with telehealth and comfort with telehealth technology as a moderator of treatment response for PTSD and depression in veterans. Int J Psychiatry Med 2014;48(1):57-67. [doi: 10.2190/PM.48.1.e] [Medline: 25354926]
204. Pyne JM, Fortney JC, Tripathi SP, Maciejewski ML, Edlund MJ, Williams DK. Cost-effectiveness analysis of a rural telemedicine collaborative care intervention for depression. Arch Gen Psychiatry 2010 Aug;67(8):812-821. [doi: 10.1001/archgenpsychiatry.2010.82] [Medline: 20679589]
205. Rabinowitz T, Murphy KM, Amour JL, Ricci MA, Caputo MP, Newhouse PA. Benefits of a telepsychiatry consultation service for rural nursing home residents. Telemed J E Health 2010;16(1):34-40. [doi: 10.1089/tmj.2009.0088] [Medline: 20070161]
206. Ransom D, Heckman TG, Anderson T, Garske J, Holroyd K, Basta T. Telephone-delivered, interpersonal psychotherapy for HIV-infected rural persons with depression: a pilot trial. Psychiatr Serv 2008 Aug;59(8):871-877. [doi: 10.1176/ps.2008.59.8.871] [Medline: 18678684]
207. Revicki DA, Tohen M, Gyulai L, Thompson C, Pike S, Davis-Vogel A, et al. Telephone versus in-person clinical and health status assessment interviews in patients with bipolar disorder. Harv Rev Psychiatry 1997;5(2):75-81. [doi: 10.3109/10673229709034730] [Medline: 9385024]
208. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 08;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
209. Robinson E, Titov N, Andrews G, McIntyre K, Schwencke G, Solley K. Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 03;5(6):e10942 [FREE Full text] [doi: 10.1371/journal.pone.0010942] [Medline: 20532167]
210. Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. Am J Psychiatry 1997 Nov;154(11):1593-1598. [doi: 10.1176/ajp.154.11.1593] [Medline: 9356570]
211. Ruskin PE, Silver-Aylaian M, Kling MA, Reed SA, Bradham DD, Hebel JR, et al. Treatment outcomes in depression: comparison of remote treatment through telepsychiatry to in-person treatment. Am J Psychiatry 2004 Aug;161(8):1471-1476. [doi: 10.1176/appi.ajp.161.8.1471] [Medline: 15285975]
212. Russell TG, Gillespie N, Hartley N, Theodoros D, Hill A, Gray L. Exploring the predictors of home telehealth uptake by elderly Australian healthcare consumers. J Telemed Telecare 2015 Dec;21(8):485-489. [doi: 10.1177/1357633X15606264] [Medline: 26391512]
213. Salfi J, Ploeg J, Black ME. Seeking to understand telephone support for dementia caregivers. West J Nurs Res 2005 Oct;27(6):701-721. [doi: 10.1177/0193945905276882] [Medline: 16157943]
214. Schutte JL, McCue MP, Parmanto B, McGonigle J, Handen B, Lewis A, et al. Usability and reliability of a remotely administered adult autism assessment, the autism diagnostic observation schedule (ADOS) module 4. Telemed J E Health 2015 Mar;21(3):176-184 [FREE Full text] [doi: 10.1089/tmj.2014.0011] [Medline: 25569603]
215. Seidel RW, Kilgus MD. Agreement between telepsychiatry assessment and face-to-face assessment for Emergency Department psychiatry patients. J Telemed Telecare 2014 Mar;20(2):59-62. [doi: 10.1177/1357633X13519902] [Medline: 24414395]
216. Shealy KM, Davidson TM, Jones AM, Lopez CM, de Arellano MA. Delivering an evidence-based mental health treatment to underserved populations using telemedicine: the case of a trauma-affected adolescent in a rural setting. Cognit Behav Pract 2015 Aug;22(3):331-344. [doi: 10.1016/j.cbpra.2014.04.007]
217. Shore JH, Savin D, Orton H, Beals J, Manson SM. Diagnostic reliability of telepsychiatry in American Indian veterans. Am J Psychiatry 2007 Jan;164(1):115-118. [doi: 10.1176/ajp.2007.164.1.115] [Medline: 17202552]
218. Shore JH, Brooks E, Anderson H, Bair B, Dailey N, Kaufmann LJ, et al. Characteristics of telemental health service use by American Indian veterans. Psychiatr Serv 2012 Feb 01;63(2):179-181 [FREE Full text] [doi: 10.1176/appi.ps.201100098] [Medline: 22302338]
219. Shore P, Goranson A, Ward MF, Lu MW. Meeting veterans where they're @: a VA Home-Based Telemental Health (HBTMH) pilot program. Int J Psychiatry Med 2014;48(1):5-17. [doi: 10.2190/PM.48.1.b] [Medline: 25354923]
220. Simon GE, Revicki D, VonKorff M. Telephone assessment of depression severity. J Psychiatr Res 1993;27(3):247-252. [doi: 10.1016/0022-3956(93)90035-z] [Medline: 8295157]
221. Simon GE, Ludman EJ, Rutter CM. Incremental benefit and cost of telephone care management and telephone psychotherapy for depression in primary care. Arch Gen Psychiatry 2009 Oct;66(10):1081-1089. [doi: 10.1001/archgenpsychiatry.2009.123] [Medline: 19805698]
222. Simpson J, Doze S, Urness D, Hailey D, Jacobs P. Evaluation of a routine telepsychiatry service. J Telemed Telecare 2001;7(2):90-98. [doi: 10.1258/1357633011936219] [Medline: 11331046]
223. Simpson J, Doze S, Urness D, Hailey D, Jacobs P. Telepsychiatry as a routine service - the perspective of the patient. J Telemed Telecare 2001;7(3):155-160. [doi: 10.1258/1357633011936318] [Medline: 11346475]

XSL•FO
RenderX

224. Simpson S, Bell L, Britton P, Mitchell D, Morrow E, Johnston A, et al. Does video therapy work? A single case series of bulimic disorders. Eur Eat Disorders Rev 2006 Jul;14(4):226-241. [doi: 10.1002/erv.686]
225. Singh SP, Arya D, Peters T. Accuracy of telepsychiatric assessment of new routine outpatient referrals. BMC Psychiatry 2007 Oct 05;7:55 [FREE Full text] [doi: 10.1186/1471-244X-7-55] [Medline: 17919329]
226. Smith AC, Stathis S, Randell A, Best D, Ryan V, Bergwever E, et al. A cost-minimization analysis of a telepaediatric mental health service for patients in rural and remote Queensland. J Telemed Telecare 2016 Dec 02;13(3_suppl):79-83. [doi: 10.1258/135763307783247239]
227. Smolenski DJ, Pruitt LD, Vuletic S, Luxton DD, Gahm G. Unobserved heterogeneity in response to treatment for depression through videoconference. Psychiatr Rehabil J 2017 Sep;40(3):303-308. [doi: 10.1037/prj0000273] [Medline: 28604014]
228. Spaniel F, Novak T, Motlova LB, Capkova J, Slovakova A, Trancik P, et al. Psychiatrist's adherence: a new factor in relapse prevention of schizophrenia. A randomized controlled study on relapse control through telemedicine system. J Psychiatr Ment Health Nurs 2015 Dec;22(10):811-820. [doi: 10.1111/jpm.12251] [Medline: 26176646]
229. Spaulding R, Belz N, DeLurgio S, Williams AR. Cost savings of telemedicine utilization for child psychiatry in a rural Kansas community. Telemed J E Health 2010 Oct;16(8):867-871. [doi: 10.1089/tmj.2010.0054] [Medline: 20925567]
230. Spence J, Titov N, Dear BF, Johnston L, Solley K, Lorian C, et al. Randomized controlled trial of internet-delivered cognitive behavioral therapy for posttraumatic stress disorder. Depress Anxiety 2011 Jul;28(7):541-550. [doi: 10.1002/da.20835] [Medline: 21721073]
231. Spence J, Titov N, Johnston L, Dear BF, Wootton B, Terides M, et al. Internet-delivered eye movement desensitization and reprocessing (iEMDR): an open trial. F1000Res 2013;2:79 [FREE Full text] [doi: 10.12688/f1000research.2-79.v2] [Medline: 24555047]
232. Spence J, Titov N, Johnston L, Jones MP, Dear BF, Solley K. Internet-based trauma-focused cognitive behavioural therapy for PTSD with and without exposure components: a randomised controlled trial. J Affect Disord 2014 Jun;162:73-80. [doi: 10.1016/j.jad.2014.03.009] [Medline: 24767009]
233. Staton-Tindall M, Havens JR, Webster JM, Leukefeld C. METelemedicine: a pilot study with rural alcohol users on community supervision. J Rural Health 2014;30(4):422-432 [FREE Full text] [doi: 10.1111/jrh.12076] [Medline: 24800699]
234. Stefan S, David D. Face-to-face counseling versus high definition holographic projection system. Efficacy and therapeutic alliance. A brief research report. J Cognit Behav Psychother 2013;13(2):299-307 [FREE Full text]
235. Steffen AM. Anger management for dementia caregivers: a preliminary study using video and telephone interventions. Behavior Therapy 2000;31(2):281-299. [doi: 10.1016/S0005-7894(00)80016-7]
236. Stevens A, Doidge N, Goldbloom D, Voore P, Farewell J. Pilot study of televideo psychiatric assessments in an underserviced community. Am J Psychiatry 1999 May;156(5):783-785. [doi: 10.1176/ajp.156.5.783] [Medline: 10327917]
237. Strachan M, Gros DF, Ruggiero KJ, Lejuez CW, Acierno R. An integrated approach to delivering exposure-based treatment for symptoms of PTSD and depression in OIF/OEF veterans: preliminary findings. Behav Ther 2012 Sep;43(3):560-569. [doi: 10.1016/j.beth.2011.03.003] [Medline: 22697444]
238. Stubbings DR, Rees CS, Roberts LD, Kane RT. Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: randomized controlled trial. J Med Internet Res 2013 Nov 19;15(11):e258 [FREE Full text] [doi: 10.2196/jmir.2564] [Medline: 24252663]
239. Swinson RP, Fergus KD, Cox BJ, Wickwire K. Efficacy of telephone-administered behavioral therapy for panic disorder with agoraphobia. Behav Res Ther 1995 May;33(4):465-469. [doi: 10.1016/0005-7967(94)00061-n] [Medline: 7755536]
240. Tan G, Teo I, Srivastava D, Smith D, Smith SL, Williams W, et al. Improving access to care for women veterans suffering from chronic pain and depression associated with trauma. Pain Med 2013 Jul;14(7):1010-1020. [doi: 10.1111/pme.12131] [Medline: 23659470]
241. Tang WK, Chiu H, Woo J, Hjelm M, Hui E. Telepsychiatry in psychogeriatric service: a pilot study. Int J Geriatr Psychiatry 2001 Jan;16(1):88-93. [doi: 10.1002/1099-1166(200101)16:1<88::aid-gps282>3.0.co;2-w] [Medline: 11180491]
242. Tarp K, Bojesen AB, Mejldal A, Nielsen AS. Effectiveness of optional videoconferencing-based treatment of alcohol use disorders: randomized controlled trial. JMIR Ment Health 2017 Sep 29;4(3):e38 [FREE Full text] [doi: 10.2196/mental.6713] [Medline: 28963093]
243. Taylor S, Thordarson DS, Spring T, Yeh AH, Corcoran KM, Eugster K, et al. Telephone-administered cognitive behavior therapy for obsessive-compulsive disorder. Cogn Behav Ther 2003;32(1):13-25. [doi: 10.1080/16506070310003639] [Medline: 16291531]
244. Théberge-Lapointe N, Marchand A, Langlois F, Gosselin P, Watts S. Efficacy of a cognitive-behavioural therapy administered by videoconference for generalized anxiety disorder. Eur Rev Appl Psychol 2015 Jan;65(1):9-17. [doi: 10.1016/j.erap.2014.10.001]
245. Thorp SR, Fidler J, Moreno L, Floto E, Agha Z. Lessons learned from studies of psychotherapy for posttraumatic stress disorder via video teleconferencing. Psychol Serv 2012 May;9(2):197-199. [doi: 10.1037/a0027057] [Medline: 22662733]
246. Titov N, Andrews G, Schwencke G, Drobny J, Einstein D. Shyness 1: distance treatment of social phobia over the internet. Aust N Z J Psychiatry 2008 Jul;42(7):585-594. [doi: 10.1080/00048670802119762] [Medline: 18612862]
247. Titov N, Andrews G, Schwencke G. Shyness 2: treating social phobia online: replication and extension. Aust N Z J Psychiatry 2008 Jul;42(7):595-605. [doi: 10.1080/00048670802119820] [Medline: 18612863]

XSL•FO
RenderX

248. Titov N, Andrews G, Choi I, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107] [Medline: 19016091]
249. Titov N, Andrews G, Robinson E, Schwencke G, Johnston L, Solley K, et al. Eclinician-assisted internet-based treatment is effective for generalized anxiety disorder: randomized controlled trial. Aust N Z J Psychiatry 2009 Jan 01;43(10):905-912. [doi: 10.1080/00048670903179269]
250. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic internet treatment for anxiety disorders: a randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]
251. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
252. Tremont G, Davis JD, Bishop DS, Fortinsky RH. Telephone-delivered psychosocial intervention reduces burden in dementia caregivers. Dementia (London) 2008;7(4):503-520 [FREE Full text] [doi: 10.1177/1471301208096632] [Medline: 20228893]
253. Tse YJ, McCarty CA, Stoep AV, Myers KM. Teletherapy delivery of caregiver behavior training for children with attention-deficit hyperactivity disorder. Telemed J E Health 2015 Jun;21(6):451-458 [FREE Full text] [doi: 10.1089/tmj.2014.0132] [Medline: 25719609]
254. Tuerk PW, Yoder M, Ruggiero KJ, Gros DF, Acierno R. A pilot study of prolonged exposure therapy for posttraumatic stress disorder delivered via telehealth technology. J Trauma Stress 2010 Feb;23(1):116-123 [FREE Full text] [doi: 10.1002/jts.20494] [Medline: 20135675]
255. Tunstall N, Prince M, Mann A. Concurrent validity of a telephone-administered version of the Gospel Oak instrument (including the SHORT-CARE). Int J Geriat Psychiatry 1997 Oct;12(10):1035-1038. [doi: 10.1002/(SICI)1099-1166(199710)12:10<1035::AID-GPS688>3.0.CO;2-9]
256. Tutty, Spangler DL, Poppleton LE, Ludman EJ, Simon GE. Evaluating the effectiveness of cognitive-behavioral teletherapy in depressed adults. Behav Ther 2010 Jun;41(2):229-236. [doi: 10.1016/j.beth.2009.03.002] [Medline: 20412887]
257. Uebelacker LA, Marootian BA, Tigue P, Haggarty R, Primack JM, Miller IW. Telephone depression care management for Latino Medicaid health plan members: a pilot randomized controlled trial. J Nerv Ment Dis 2011 Sep;199(9):678-683. [doi: 10.1097/NMD.0b013e318229d100] [Medline: 21878782]
258. Vahia IV, Ng B, Camacho A, Cardenas V, Cherner M, Depp CA, et al. Telepsychiatry for neurocognitive testing in older rural Latino adults. Am J Geriatr Psychiatry 2015 Jul;23(7):666-670 [FREE Full text] [doi: 10.1016/j.jagp.2014.08.006] [Medline: 25708655]
259. van Ballegooijen W, Riper H, Klein B, Ebert DD, Kramer J, Meulenbeek P, et al. An internet-based guided self-help intervention for panic symptoms: randomized controlled trial. J Med Internet Res 2013 Jul 29;15(7):e154 [FREE Full text] [doi: 10.2196/jmir.2362] [Medline: 23896222]
260. van Bastelaar KM, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
261. Ward-King J, Cohen IL, Penning H, Holden JJ. Brief report: telephone administration of the autism diagnostic interview - revised: reliability and suitability for use in research. J Autism Dev Disord 2010 Oct;40(10):1285-1290. [doi: 10.1007/s10803-010-0987-x] [Medline: 20195733]
262. Watson CG, Anderson PE, Thomas D, Nyberg K. Comparability of telephone and face to face diagnostic interview schedules. J Nerv Ment Dis 1992 Aug;180(8):534-535. [doi: 10.1097/00005053-199208000-00010] [Medline: 1500937]
263. Wells KB, Burnam MA, Leake B, Robins LN. Agreement between face-to-face and telephone-administered versions of the depression section of the NIMH Diagnostic Interview Schedule. J Psychiatr Res 1988;22(3):207-220. [doi: 10.1016/0022-3956(88)90006-4] [Medline: 3225790]
264. Whealin JM, Seibert-Hatalsky LA, Howell JW, Tsai J. E-mental health preferences of Veterans with and without probable posttraumatic stress disorder. J Rehabil Res Dev 2015;52(6):725-738 [FREE Full text] [doi: 10.1682/JRRD.2014.04.0113] [Medline: 26562090]
265. Wierwille JL, Pukay-Martin ND, Chard KM, Klump MC. Effectiveness of PTSD telehealth treatment in a VA clinical sample. Psychol Serv 2016 Nov;13(4):373-379. [doi: 10.1037/ser0000106] [Medline: 27657798]
266. Wilz G, Schinköthe D, Soellner R. Goal attainment and treatment compliance in a cognitive-behavioral telephone intervention for family caregivers of persons with dementia. GeroPsych 2011 Sep;24(3):115-125. [doi: 10.1024/1662-9647/a000043]
267. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: a randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]
268. Winter L, Gitlin LN. Evaluation of a telephone-based support group intervention for female caregivers of community-dwelling individuals with dementia. Am J Alzheimers Dis Other Demen 2006;21(6):391-397 [FREE Full text] [doi: 10.1177/1533317506291371] [Medline: 17267370]
269. Wray LO, Shulan MD, Toseland RW, Freeman KE, Vásquez BE, Gao J. The effect of telephone support groups on costs of care for veterans with dementia. Gerontologist 2010 Oct;50(5):623-631. [doi: 10.1093/geront/gnq040] [Medline: 20507926]

XSL•FO
RenderX

270. Yeung A, Johnson DP, Trinh NH, Weng WC, Kvedar J, Fava M. Feasibility and effectiveness of telepsychiatry services for chinese immigrants in a nursing home. Telemed J E Health 2009 May;15(4):336-341. [doi: 10.1089/tmj.2008.0138] [Medline: 19441951]
271. Yoshino A, Shigemura J, Kobayashi Y, Nomura S, Shishikura K, Den R, et al. Telepsychiatry: assessment of televideo psychiatric interview reliability with present- and next-generation internet infrastructures. Acta Psychiatr Scand 2001 Sep;104(3):223-226. [doi: 10.1034/j.1600-0447.2001.00236.x] [Medline: 11531660]
272. Yuen EK, Herbert JD, Forman EM, Goetter EM, Juarascio AS, Rabin S, et al. Acceptance based behavior therapy for social anxiety disorder through videoconferencing. J Anxiety Disord 2013 May;27(4):389-397. [doi: 10.1016/j.janxdis.2013.03.002] [Medline: 23764124]
273. Yuen EK, Gros DF, Price M, Zeigler S, Tuerk PW, Foa EB, et al. Randomized controlled trial of home-based telehealth versus in-person prolonged exposure for combat-related PSTD in Veterans: preliminary results. J Clin Psychol 2015 Jun;71(6):500-512. [doi: 10.1002/jclp.22168] [Medline: 25809565]
274. Zheng P, Gray MJ. Telehealth-based therapy connecting rural mandarin-speaking traumatized clients with a mandarin-speaking therapist. Clin Case Stud 2014 Apr 13;13(6):514-527. [doi: 10.1177/1534650114529182]
275. Zheng W, Nickasch M, Lander L, Wen S, Xiao M, Marshalek P, et al. Treatment outcome comparison between telepsychiatry and face-to-face buprenorphine medication-assisted treatment for opioid use disorder: a 2-year retrospective data analysis. J Addict Med 2017;11(2):138-144 [FREE Full text] [doi: 10.1097/ADM.0000000000000287] [Medline: 28107210]
276. Ziemba SJ, Bradley NS, Landry LA, Roth CH, Porter LS, Cuyler RN. Posttraumatic stress disorder treatment for Operation Enduring Freedom/Operation Iraqi Freedom combat veterans through a civilian community-based telemedicine network. Telemed J E Health 2014 May;20(5):446-450. [doi: 10.1089/tmj.2013.0312] [Medline: 24617961]
277. Aafjes-van Doorn K, Békés V, Prout TA. Grappling with our therapeutic relationship and professional self-doubt during COVID-19: will we use video therapy again? Counsel Psychol Q 2020 Jun 03:1-12. [doi: 10.1080/09515070.2020.1773404]
278. Guidelines for the practice of telepsychology. American Psychological Association. 2013. URL: https://www.apa.org/practice/guidelines/telepsychology [accessed 2021-06-26]
279. Hoffman L, Wisniewski H, Hays R, Henson P, Vaidyam A, Hendel V, et al. Digital Opportunities for Outcomes in Recovery Services (DOORS). J Psychiatr Pract 2020;26(2):80-88. [doi: 10.1097/pra.0000000000000450]
280. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
281. Richardson S, Carroll C, Close J, Gordon A, O'Brien J, Quinn T, et al. Research with older people in a world with COVID-19: identification of current and future priorities, challenges and opportunities. Age Ageing 2020 Oct 23;49(6):901-906 [FREE Full text] [doi: 10.1093/ageing/afaa149] [Medline: 32584954]
282. Papatheodorou S. Umbrella reviews: what they are and why we need them. Eur J Epidemiol 2019 Jun 9;34(6):543-546. [doi: 10.1007/s10654-019-00505-6] [Medline: 30852716]
283. Mulder R, Singh AB, Hamilton A, Das P, Outhred T, Morris G, et al. The limitations of using randomised controlled trials as a basis for developing treatment guidelines. Evid Based Ment Health 2018 Feb 14;21(1):4-6. [doi: 10.1136/eb-2017-102701] [Medline: 28710065]
284. Johns B, Baltussen R, Hutubessy R. Programme costs in the economic evaluation of health interventions. Cost Eff Resour Alloc 2003 Feb 26;1(1):1 [FREE Full text] [doi: 10.1186/1478-7547-1-1] [Medline: 12773220]

Abbreviations

AMSTAR: Assessment of Multiple Systematic Reviews
ARC: Applied Research Collaboration
BAME: Black, Asian, and minority ethnic
NHS: National Health Service
NIHR: National Institute for Health Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder

XSL•FO
RenderX

Edited by R Kukafka; submitted 14.12.20; peer-reviewed by A Khoja, L Le; comments to author 12.03.21; revised version received 30.03.21; accepted 30.04.21; published 20.07.21.

Please cite as:
Barnett P, Goulding L, Casetta C, Jordan H, Sheridan-Rains L, Steare T, Williams J, Wood L, Gaughran F, Johnson S
Implementation of Telemental Health Services Before COVID-19: Rapid Umbrella Review of Systematic Reviews
J Med Internet Res 2021;23(7):e26492
URL: https://www.jmir.org/2021/7/e26492
doi:10.2196/26492
PMID:34061758

©Phoebe Barnett, Lucy Goulding, Cecilia Casetta, Harriet Jordan, Luke Sheridan-Rains, Thomas Steare, Julie Williams, Lisa Wood, Fiona Gaughran, Sonia Johnson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Exploring the Role of Web-Based Interventions in the Self-management of Dementia: Systematic Review and Narrative Synthesis

Abigail Rebecca Lee[1], BSc, MSc; Esther Vera Gerritzen[1], BSc, MSc; Orii McDermott[1], ARCM, BA, MMT, PhD; Martin Orrell[1], BMedSci, MBBS, PhD

Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Abigail Rebecca Lee, BSc, MSc
Mental Health and Clinical Neurosciences
School of Medicine
University of Nottingham
Jubilee Campus
Triumph Road
Nottingham
United Kingdom
Phone: 44 7890 021703
Email: abigail.lee1@nottingham.ac.uk

Abstract

Background: The increasing prevalence of dementia has promoted a move toward equipping people with the skills required for greater self-management of the condition to enable a better quality of life. Self-management encompasses numerous skills, such as goal setting and decision making, which aim to improve an individual's physical and mental well-being when they live with long-term health conditions. Effective self-management may lead to increased well-being and quality of life. Reviews of web-based and app-based interventions have suggested that they have the potential to provide self-management support for people living with a range of conditions, including dementia.

Objective: The aim of this review is to explore the existing use of web-based or app-based interventions that facilitate or support self-management in dementia and discuss their effectiveness in promoting self-management and independence.

Methods: A total of 5 electronic databases were systematically searched for relevant articles published between January 2010 and March 2020. Included studies were appraised using the Downs and Black checklist and the Critical Appraisal Skills Program qualitative research checklist. A narrative synthesis framework was applied using tables and conceptual mapping to explore the relationships within and among studies.

Results: A total of 2561 articles were identified from the initial search, of which 11 (0.43%) met the inclusion criteria for the final analysis. These included 5 quantitative, 4 mixed methods, and 2 qualitative studies. All the included articles were of fair to high quality across the two appraisal measures. Interventions were delivered through a range of web-based and app-based technologies and targeted several self-management concepts. However, there was inconsistency regarding the domains, often affected by dementia, that were targeted by the interventions reviewed.

Conclusions: Web-based and app-based interventions for dementia can be delivered through a range of means and can target different aspects of self-management. The small number of studies included in this review report positive outcomes that seem to support the use of these interventions for people living with dementia. However, there is a clear need for more high-quality research into this type of intervention delivery and for studies that use a much larger number of participants across the dementia spectrum. Future research should consider the barriers to and facilitators of intervention adoption highlighted in this review and whether interventions can encompass the physical, social, cognitive, and emotional domains affected by dementia.

(J Med Internet Res 2021;23(7):e26551) doi:10.2196/26551

KEYWORDS

systematic review; narrative synthesis; dementia; web; app; online; mobile phone; self-management

XSL•FO
RenderX

Introduction

Background

Technology-based interventions have the potential to provide practical and effective delivery of support to affected populations across a range of health conditions [1]. Their role in dementia care is still emerging, and more research is needed to explore their current use and potential impact and highlight gaps in the literature and knowledge [2].

Dementia currently affects an estimated 885,000 people in the United Kingdom [3,4] and leads to impaired ability and performance across multiple cognitive domains, such as memory, cognitive ability, and communication, which appear even in the early stages [4]. The provision of care for people aged above 65 years who are living with dementia currently costs the United Kingdom £34.7 (US $47.9) billion a year and is expected to rise to £94.1 (US $129.9) billion by 2040 [5]. With the ever-increasing aging population, it is estimated that 1.6 million people will have a dementia diagnosis by 2040 [5]. Social care costs alone amount to £15.7 (US $21.6) billion, and the hours of unpaid care by families equate to £13.9 (US $19.2) billion a year [5]. Enabling people living with dementia to manage their condition more effectively, improve their overall well-being, and maintain their independence for as long as possible may provide benefits for both the population living with dementia and the health and social care sectors [6]. The role of technology-based interventions in dementia care is still emerging; however, they may offer the potential to provide practical and effective delivery of support for people living with dementia and their families [6].

The UK government highlighted the importance of enabling people with dementia to live well and independently in their dementia action plan [7]. Self-management was identified as a potential strategy in response to the increasing incidence and prevalence of dementia and in helping people and their families to retain control over their lives. Self-management encompasses multiple components that can support an individual to improve their physical and mental well-being, either independently or in collaboration with their health care team [8]. These components include goal setting, decision making, problem solving, accessing and using resources, strong collaboration between patients and health professionals, and patient activation [8-10]. The latter refers to the knowledge, skills, and confidence an individual has in managing their long-term condition and overall health and has been linked to a lower number of medical appointments and hospital admissions [11].

The lived experiences of people with dementia vary considerably, and it has been suggested that this may be due to the interaction between cognitive impairment and a range of psychological and social factors [12]. One review found that multicomponent, nonpharmacological interventions for people living with dementia had a positive effect on the activities of daily living, cognitive functioning, and mood [13]. In addition, interventions targeted at dyads were found to have positive effects on the quality of life of people with dementia and their caregivers. Oyebode and Parveen [12] extended the previous evidence by updating the evidence base to consider randomized controlled trials (RCTs), controlled studies, and reviews from 2008 to 2015. The 61 studies and reviews included covered the entire dementia care pathway, from community-dwelling people to residential care and end-of-life care, and considered interventions aimed at caregivers [12]. Many of the publications included discussed residential care, with a focus on managing the behavioral symptoms of dementia. The authors concluded that more research was needed into care within the community-dwelling dementia population and a greater focus on interventions that help to enrich the overall quality of life.

A review of web-based interventions that targeted support and education to informal caregivers found that they have potential benefits for both the supporter and the person with dementia [14]. A systematic search of the literature pertaining to RCTs of web-based interventions resulted in 17 studies. Interventions were found to be effective in decreasing symptoms of depression and anxiety in informal caregivers but failed to significantly reduce caregiver burden or improve quality of life. However, 6 studies demonstrated that caregiver interventions had the potential to positively improve the symptoms of depression and anxiety in caregivers and the quality of life of people with dementia. The review suggested that, when tailored to individuals and targeted at both caregivers and people with dementia, web-based interventions have the potential to improve the well-being and quality of life of all involved in informal dementia care.

Although the older population is generally perceived to have fewer technology skills, there is an emerging evidence base suggesting that technology plays a role in the self-management of dementia. In fact, it has been suggested that technology has five potential roles in dementia care [15]: facilitating declining cognition, enabling better performance of daily activities, ensuring safety, helping maintain active social involvement, and providing support and reassurance for informal caregivers. All these roles aim to assist people living with dementia to maintain their independence, improve their quality of life, and contribute to their self-management.

Research focused specifically on app-based interventions targeted at people living with dementia has also supported their use in the self-management of the condition. A study exploring the use of tablet computers and apps by people with mild dementia demonstrated that people were quickly able to learn how to use new technology and engage positively with the content of the apps [16]. The findings highlighted the importance of motivational benefits for people to incorporate new technology into their daily lives, such as improving their self-management and quality of life. Access to informal technology support to aid adoption was shown to be valued by people living with dementia and their families. However, consideration should be given to individualizing interventions to encourage engagement [16]. Several factors should be considered when creating and delivering self-management interventions (SMIs) in dementia to maximize their potential benefit and use.

Dementia is a chronic, progressive condition that affects multiple faculties in daily life [4]. The evidence base for self-management in dementia is limited, particularly regarding support for people

XSL•FO
RenderX

living with mild dementia [8,17]. Therefore, an in-depth review of the current knowledge and use of interventions, particularly regarding the role of technology, is needed.

Objectives

There are a range of nonpharmacological digital interventions that may be beneficial to people living with dementia, such as cognitive stimulation therapy [18]. However, the aim of this review is to explore the existing use of web- or app-based interventions that facilitate or support self-management in dementia, the concepts they target, and their effectiveness.

The findings are likely to be useful to health services and policy makers when considering how to include self-management in dementia and to researchers to help design better studies on the effectiveness of web- and app-based SMIs. This review could provide useful insights into the role of web- and app-based interventions in the self-management of dementia, and the findings should be considered in clinical practice. A protocol was written for this review but was not registered with PROSPERO [19].

Methods

Overview

Narrative synthesis is one approach to the systematic review and synthesis of findings from multiple studies and different methodologies. Although it allows for the inclusion of statistical data, the distinguishing characteristic of narrative synthesis is the use of a textual approach to summarize and describe findings to form a story from the included studies.

Search Strategy

A systematic search was conducted across five electronic databases in February 2020: Cochrane (Central Register of Controlled Trials), Web of Science, PubMed, Scopus, and ProQuest (Science Database, Technology Collection, PsycArticles, and Social Science Database). After scoping the literature, a trial-and-error process was applied to explore search term combinations. With each combination, every third title and abstract were screened on the first two pages of results to determine whether they were relevant to the review questions. The key terms found were combined to create the final search: (web* OR online* OR computer* OR internet* OR app* OR smartphone*) AND (intervention* OR support*) AND (self-manag* OR independ*) AND (dement*). Independence was found to be a term often used in discussions about self-management; therefore, it was included in the final search. Terms such as *tablet* were excluded from the search because of their connotations with pharmacological interventions found during the initial scope of the literature. The search included research, journal, and review or evaluation articles, as it was thought that these would encompass novel research and evaluation studies. The date limits of January 2010 to March 2020 were placed on the search to encompass any prospective publications.

Study Selection

The search results were imported into EndNote (Clarivate Analytics), and duplicates were removed. Each title and abstract were read twice and vetted by the primary reviewer (ARL), with the inclusion criteria acting as a guide to identify possible papers. A second reviewer (EVG) independently examined 5% of the total results to provide a consensus on the quality of the search. Potentially relevant references were imported into Rayyan [20], ready for a full-text review by the 2 reviewers (ARL and EVG). Each reviewer independently read the full texts twice before deciding whether to include or exclude the review. Any conflicts regarding the inclusion or exclusion of papers at any stage of the process were discussed by the two reviewers. A manual search of the references from the included papers was conducted for any suitable additions.

Inclusion Criteria

The inclusion criteria were as follows:

- Participant population included adults aged 18 years or older, with a confirmed diagnosis of dementia.
- Participant population was community dwelling.
- Included a web- or app-based intervention aimed at improving self-management or independence for people living with dementia.
- Intervention was for independent or dyadic use (involvement from an informal supporter).
- Included RCTs or quasi-experimental, observational, qualitative, or mixed methods studies.
- Publication dates were between January 2010 and March 2020. These years were selected based on the definition of web-based interventions by Barak et al [21].

Exclusion Criteria

The exclusion criteria were as follows:

- Protocol papers, opinion pieces, conference abstracts, scoping reviews, or systematic reviews.
- Interventions that were exclusively for supporters.
- Studies with a focus on care management and community-delivered interventions where the planning and coordination of dementia care was the focus [22].
- Published in a language other than English, and a translation was not available.

Data Extraction

The principal reviewer (ARL) completed the data extraction using bespoke extraction forms based on the guidance of the Centre for Reviews and Dissemination for systematic reviews [23]. The data extraction forms were piloted before the review. A second independent review of the completed data extraction was provided by EVG. The following data items were extracted: (1) study information, (2) study characteristics, (3) population characteristics, (4) intervention, (5) outcome data, and (6) results.

Quality Assessment and Risk of Bias in Individual Studies

The quality of studies assessed aspects such as the appropriateness of the study design, the potential risk of bias, and the quality of reporting. A total of 2 assessment tools were used: the modified Downs and Black [24] checklist, as used in Trac et al [25], to measure study quality for quantitative trials and the Critical Appraisal Skills Programme (CASP) checklist

XSL•FO
RenderX

for qualitative research [26]. Mixed methods studies were assessed using both checklists.

The scoring system used for the modified Downs and Black checklist followed that outlined in the study by O'Connor et al [27], with 24-28 points regarded as *excellent*, 19-23 as *good*, 14-18 as *fair*, and less than 14 as *poor*. The 10-item CASP checklist had three response options: *meeting the criteria*, *unable to tell*, and *not meeting the criteria*. It was scored according to the method detailed in Stansfeld et al [28], with *meeting the criteria* given a score of 1 and *unable to tell* or *not meeting the criteria* given a score of 0. For the tenth item, which asks how valuable the research is and does not provide the response options, the principal reviewer decided whether to award a score of 1. The principal reviewer administered a scoring system in which a score of 4 or less was defined as *poor*, 5-7 as *moderate*, and 8 or above as *high*. These tools were selected as they are suitable for randomized, nonrandomized, and qualitative studies. They have also been used in previous narrative synthesis systematic reviews [28,29] and are recommended by the Centre for Reviews and Dissemination [23].

Data Synthesis

Narrative synthesis allows for the inclusion of qualitative, quantitative, and mixed methods studies and for a systematic yet transparent review of results. Therefore, owing to the diverse selection of studies and review transparency, narrative synthesis was viewed as the most suitable option for this review.

Unlike more analytical approaches to literature reviews, such as meta-analyses, narrative synthesis does not rely on a rigorously tested structured technique [30]. Popay et al [30] created guidance and a framework of four interconnecting elements to improve the transparency of narrative synthesis reviews. This review applied the following guidance and framework:

- Developed a theory of how the intervention works, why, and for whom: a scoping of the relevant literature provided a greater understanding of the review topic, and the rationale for using web- or app-based interventions in dementia studies was considered. This stage guided the research questions, development of the search terms, and inclusion criteria for the review.
- Developed a preliminary synthesis of findings of included studies: data were extracted from each of the studies and tabulated. Descriptive summaries of the same features from each study were extracted and tabulated to help with the initial comparison. Studies were clustered according to the methodology: quantitative, qualitative, and mixed methods.
- Explored relationships within and between studies: concept mapping was used on the extracted data on study interventions to explore the similarities and differences between the studies and the factors that might have affected this.
- Assessed the robustness of the synthesis: two validity assessment tools were used in this study. Quantitative studies were assessed using the modified Downs and Black checklist [25], and qualitative studies were assessed using the CASP checklist [26]. Studies with mixed methodologies were assessed using both tools.

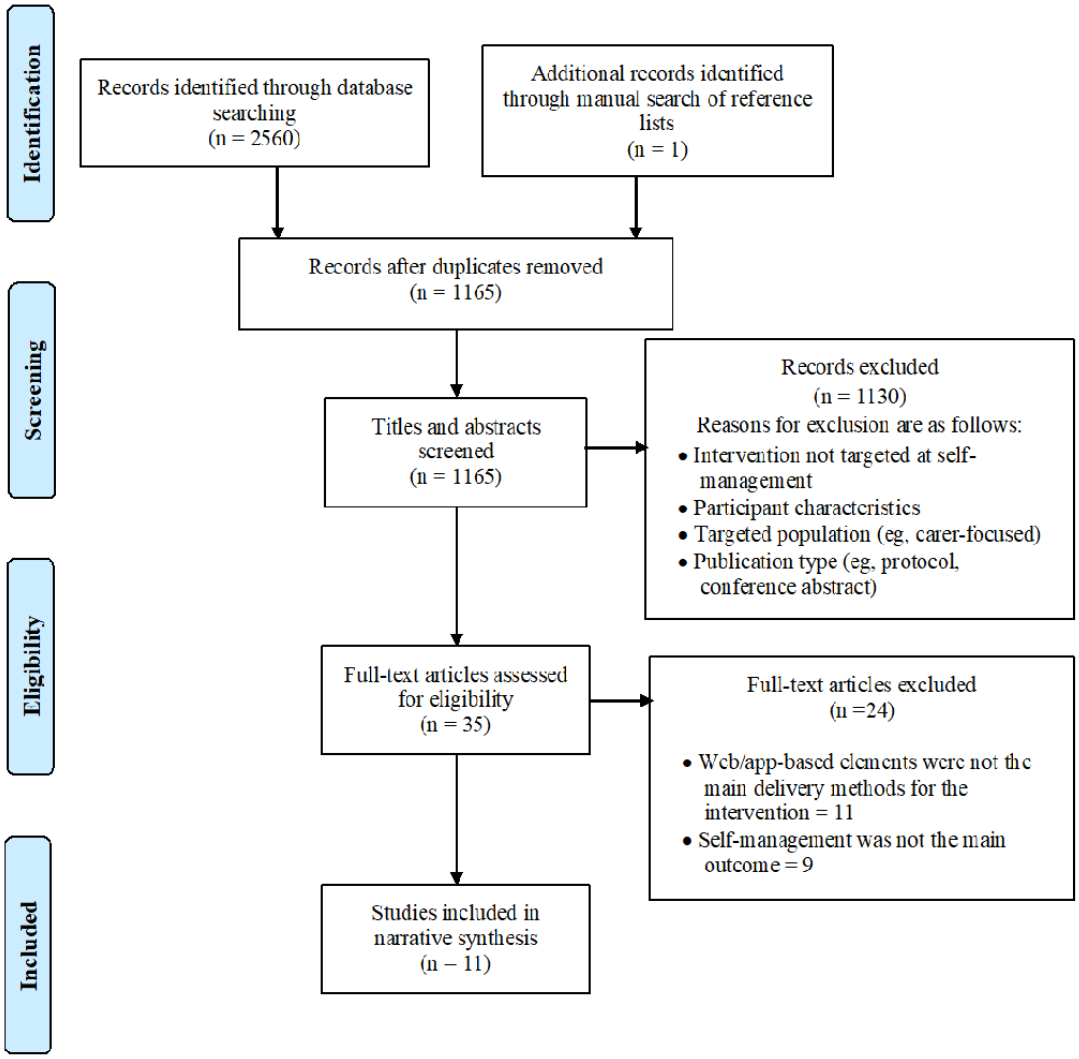
Results

Reviewing Process

A total of 2560 references were identified using the search strategy. After duplicates were removed, 1164 references remained, and their titles and abstracts were screened for inclusion criteria. Of these, 1130 were excluded as they did not focus on relevant interventions or include participants with dementia, leaving 34 papers for full-text screening. One additional paper was found through a manual search of the reference lists of the papers selected for full-text screening. After a full-text review conducted by the principal and secondary reviewers, 11 papers met the inclusion criteria and were accepted for this review. The main reasons for exclusion were that the app- or web-based interventions were not the primary focus of the study; they were not described in sufficient detail for analysis, for example, lacking description of the intervention and mode of delivery; and the outcome measures were not relevant to self-management in relation to independence. Figure 1 shows a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) diagram of the study selection process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) diagram of the search and review process.



Preliminary Synthesis of Findings

Study Characteristics

The included articles varied in location (Denmark=2, Sweden=1, United Kingdom=1, Netherlands=1, and United States=1); however, several studies did not specify a country (n=5). For studies with an unspecified location, the primary reviewer contacted the corresponding author but received no reply. Day and activity centers were the most common locations for the interventions (n=6), with private homes being the second most popular (n=5). Almost all the articles had either quantitative (n=5) or mixed methodology (n=4), and a nonrandomized, nonconcurrent multiple baseline approach was the most common study design (n=6). This meant that data from multiple baseline and intervention sessions were not collected simultaneously for all participants. No control groups or blinding procedures were used in any of the included studies.

Participant Characteristics

A total of 189 people living with dementia participated across the included studies, with an age range of 59-92 years. All studies had small sample sizes of ≤11, except for one study, which had 116 participants [31]. Alzheimer disease was the most common diagnosis among study participants (n=7), and the Mini Mental State Exam score was the most common measure used to describe participants (n=8). Scores varied from <6 to 22, indicating that the participants had mild to severe dementia. Participants were mainly recruited from day and activity centers for people living with dementia (n=6) or from memory clinics (n=3). Supporters were recruited in 4 studies, 2 as part of a dyad [32,33], and 2 as supporters [31,34]. Of the 121 supporters recruited, 119 (98.3%) were informal and 2 (1.7%) were formal (see Tables S1 and S2 of Multimedia Appendix 1 [31-41] and Multimedia Appendix 2 [31-41], respectively, for further details of the study and participant characteristics).

Exploring Relationships Within and Among Studies

Robustness of the Synthesis

The quality of the included studies varied between fair and high. All the quantitative studies [35-39] were of fair quality, in accordance with the Downs and Black checklist scoring. These studies scored highly on reporting aims, intervention details, measuring outcome measures and providing a comprehensive summary of their findings. Mixed methods studies [31-33,40] scored high or moderate on the qualitative CASP checklist but fair on the quantitative checklist. Qualitative commentary on participant recruitment and the summary of findings

XSL•FO
RenderX

complemented the quantitative reporting of aims, outcome measures, intervention details, and participant numbers and characteristics. Both measures suggested that greater reporting of data analyses, ethical considerations, the acknowledgment of monitoring for adverse events, and the inclusion of blinding would strengthen the methodology and study reporting. Of the 2 qualitative studies, one [34] scored highly on the CASP, whereas the other was moderate [41]. The reporting of study aims, data collection, and findings was strong; however, more details on the data analysis techniques used, reasoning for the chosen research design, and the relationship between researchers and participants would have been preferred. In addition, wider contribution of the research could have been discussed more thoroughly in both papers. The content of the included studies was judged to be of sufficient quality and robust enough to be included in the narrative synthesis. Table 1 shows the quality assessment scores of each of the included studies.

Table 1. Quality assessment scores.

Study	Methodology	Quality assessment score		Quality
		Value, n (%)	Total, N	
Perilli et al (2012) [35]	Quantitative	15 (54)	28	Fair
Perilli et al (2013) [36]	Quantitative	15 (54)	28	Fair
Lancioni et al (2017) [37]	Quantitative	16 (57)	28	Fair
Lancioni et al (2018) [38]	Quantitative	14 (50)	28	Fair
Lancioni et al (2019) [39]	Quantitative	14 (50)	28	Fair
Thorpe et al (2019) [32]	Mixed methods (quantitative)	14 (50)	28	Fair
Thorpe et al (2019) [32]	Mixed methods (qualitative)	7 (70)	10	Moderate
Øksnebjerg et al (2020) [31]	Mixed methods (quantitative)	14 (50)	28	Fair
Øksnebjerg et al (2020) [31]	Mixed methods (qualitative)	9 (90)	10	High
Kerssens et al (2015) [33]	Mixed methods (quantitative)	15 (54)	28	Fair
Kerssens et al (2015) [33]	Mixed methods (qualitative)	6 (60)	10	Moderate
McGoldrick et al (2019) [40]	Mixed methods (quantitative)	17 (61)	28	Fair
McGoldrick et al (2019) [40]	Mixed methods (qualitative)	9 (90)	10	High
Kerkhof et al (2019) [34]	Qualitative	8 (80)	10	High
Boman et al (2014) [41]	Qualitative	7 (70)	10	Moderate

Interventions

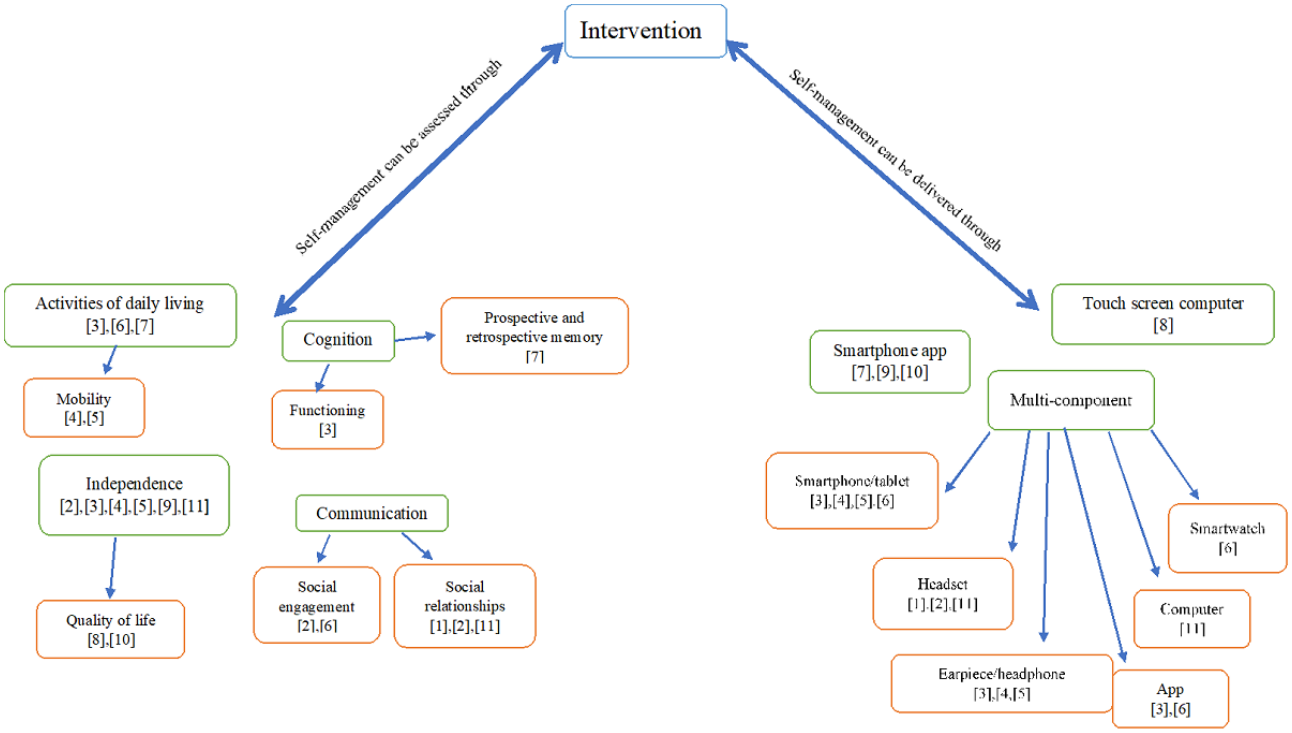
Concept mapping enabled a clear comparison of the interventions among the included studies. All the studies described their interventions in detail. There was a range of web- and app-based technologies used to deliver SMIs: touch screen computers (n=1), smartphone apps (n=3), and multicomponent (n=7). Of the multicomponent interventions, smartphones or tablets were the most commonly used (n=4), followed by earpieces or headphones (n=3) and headsets (n=3), although apps (n=2), computers (n=1), and smartwatches (n=1) were also used. These findings suggest that apps are becoming more popular in the delivery of interventions, either alone or as part of a more complex, multicomponent method. A total of 2 studies examined the same intervention but with different participants [35,36], and 3 others focused on a similar alternative intervention in different participant groups [37-39].

There were similarities and differences among the aims of the studies with regard to the self-management concepts targeted by the interventions. In total, 7 of the studies focused on interventions that targeted more than one self-management concept, although no intervention covered all domains or self-management concepts. One study targeted three concepts, 6 studies considered two concepts, and 3 studies focused on one self-management concept. Four overarching self-management concepts were widely assessed across the included studies: independence, activities of daily living, communication, and cognition.

Independence was the most commonly identified concept (n=8). A total of 2 studies focused on the effect of independence on the quality of life. Other popular concepts targeted by interventions in several studies were improving activities of daily living (n=5) and communication (n=5). Studies that explored communication could be divided into enhancing social relationships (n=3) and promoting social engagement (n=2). A total of 2 studies centered on improving cognitive functioning and memory enhancement. Figure 2 shows the intervention concept map. The numbers refer to the study identities found in the study and outcome tables.

XSL•FO
RenderX

Figure 2. Concept map of self-management interventions. Numbers indicate the study reference number in tables.



All interventions could be tailored or modified in their delivery to fit individual needs or goals. The study period was reported either through the number of intervention sessions or the number of days, with one exception where the duration of the session was provided. The number of intervention sessions varied between 20 and 119 sessions and the number of days from 24 days to 9 months. Researchers or the research team was heavily involved in the intervention setup and provision across all 11 included studies.

Outcomes

Table S3 (Multimedia Appendix 3 [31-41]) outlines the outcomes and key findings of each of the included studies.

Activities

A total of 6 studies focused on outcomes that measured or explored the completion of activity. In total, 2 studies measured the completion rate of independent phone calls to people who were relevant to the participants [35,36]. The mean number of independent calls in the baseline of both studies was 0; however, this increased to around 4 during the respective interventions. A similar study explored the experiences of using a mock-up videophone [41]. Observations and qualitative feedback from participants showed that they initially struggled with the new intervention but could use it independently following guidance from the research team. Participants reported that the intervention was enjoyable to use, but they would have preferred more options to individualize it. A total of 3 studies had outcomes that measured independent ambulation and object use [37-39]. The interventions in these studies appeared to have a considerable impact on participants' ability to start and complete independent activities successfully, such as making a cup of coffee or preparing food. In particular, one study reported a significant improvement in all participants executing the correct steps to complete their activities [38].

Engagement

The outcomes of the other 5 studies [31-34,40] explored the wider impact and experiences of app-based and wearable technology in dementia care. Interventions in 2 studies led to increased activity levels and a sense of independence in participants, which promoted positive engagement with daily activities [32,33]. Several issues regarding the incorporation of web- and app-based interventions in dementia were highlighted in the qualitative outcomes. Contextual and personal factors, such as a lack of confidence in using technology, concerns about dealing with technical difficulties, and forgetting to use apps, were some of the issues raised by participants and their families. These factors were key to nonadoption in the respective studies and should be considered when designing and delivering future studies in dementia care.

Adoption and Usability

The adoption and usability of apps were measured in 2 studies. One study [31] trialed the Rehabilitation in Alzheimer Disease Using Cognitive Support Technology app-based intervention, designed to assist with memory symptoms and daily activities. The overall mean Usefulness, Satisfaction, and Ease of use Questionnaire for dementia scores in this study of 40 for participants and 34 for supporters out of a total of 60 indicated a moderately high-level satisfaction rating of the intervention regarding usefulness, satisfaction, and ease of use. The researchers divided participants into adopters and nonadopters, depending on their usage of the intervention. There were 18 participants and 7 supporters who continued to use the app after the 90-day study period and were classed as *adopters*. However, 47 participants and 78 supporters did not activate the app. The survey, which was completed by 35 participants, showed that those who adopted the app were not significantly different from nonadopters in their skills, level of experience, and need for

XSL•FO
RenderX

help when using a tablet [31]. For those who did not activate or continue to use Rehabilitation in Alzheimer Disease Using Cognitive Support Technology, several reasons were given, including that it was not relevant for the stage of their condition and a preference for using nontechnology-based solutions [31].

Another study [40] used the Unified Theory of Acceptance and Use of Technology Questionnaire to assess changes in attitudes toward the use of their reminder app in eight domains. Unified Theory of Acceptance and Use of Technology Questionnaire scores were collected from 2 participants, with one showing a positive decrease in pre- and postscores but the other showed a negative increase in half of the domains. The adoption of web- and app-based interventions appears to be dependent upon individuals connecting with the intervention and feeling confident about using it and may or may not result in a positive research outcome.

Discussion

Principal Findings

After reviewing the current evidence, it is clear that web- and app-based interventions have the potential to benefit the lives and care of people living with dementia. This narrative synthesis review examined the literature discussing the use of web- and app-based technology in delivering SMIs in dementia care. From the 11 studies that met the inclusion criteria, it is apparent that a range of methodologies have been applied when researching this topic. All the included studies were generally of fair to good quality, and the results were consistent and coherent, which suggests that the synthesis was robust. However, the scores from the quality appraisal measures suggest that there is a lack of high-quality research on web- and app-based interventions. More details on participant recruitment methods and the acknowledgment of potential adverse events were needed, and the blinding of those conducting outcome measures would have strengthened the methodology. The interventions reviewed targeted independence, communication, and activities of daily living, and 7 studies focused on multiple concepts of self-management. However, there was inconsistency regarding the number of domains related to dementia self-management, such as daily living activities, that were targeted by each intervention.

Most studies had very small participant numbers, ranging from 3 to 11, except for Øksnebjerg et al [31], who recruited 116 participants living with dementia and 98 supporters. Owing to the small sample sizes, studies were unable to conduct comprehensive analyses on their results and often relied on reporting changes in the mean scores of outcome measures. Recruitment methods across studies were open to bias, as they usually relied on people who had contact with memory clinics or day centers. Therefore, the participants might not have been representative of the wider dementia population. There were no suitable RCTs, and none of the included studies reported blinding participants or researchers. This highlights the shortfall in comprehensive, large-scale RCTs of web- and app-based SMIs in dementia and identifies an area for future research.

Critical Reflection

Reflection was undertaken by the authors throughout the review process to identify any limitations or biases that could influence the review findings. As critical reflection is not a linear process, the authors acknowledge that there may be additional missed limitations. One strength of this review is that the search terms were created according to the scope of the relevant literature. This helped ensure that the final search would find the most relevant results and that the number of missed articles would be significantly reduced. Another strength is that the articles were differentiated and excluded using a standardized definition of care management. Having a definition meant a uniform exclusion of articles and a greater inclusion of self-management–focused results.

Although the search terms appear robust and the results were excluded in a uniform manner, this review has several limitations. First, the included articles were limited to those published in English or those that had an English translation available, which might have led to some relevant research being missed. It was decided to restrict participant populations to people living with dementia in the community, rather than those in residential homes or institutionalized care, which means the search strategy missed any web- or app-based SMIs in those settings. This could be a potential area for future reviews. Finally, owing to the small number of participants involved across the included studies, it is difficult for this review to provide a comprehensive evaluation of the effects of web- or app-based interventions on the self-management of dementia. There is a need for studies to explore these interventions in larger samples of people living with dementia and across a range of dementias and severities, for more significant conclusions to be drawn. As narrative synthesis takes a textual approach to analyzing evidence, the quality of methodological reporting could have biased the findings.

Comparison With Previous Work

To our knowledge, this is the first review to systematically synthesize evidence concerning web- and app-based SMIs for people living with dementia. However, previous reviews have identified digital interventions aimed at people living with noncommunicable diseases, such as cardiovascular diseases. One such review examined the potential role of digital interventions in promoting healthy behavior change and improving self-management [1]. A search of 9 databases resulted in 29 publications meeting the inclusion criteria, with these studies covering 7 different interventions. All 7 interventions were identified as web-based, with 4 also having mobile-based delivery and targeted health behaviors such as physical activity and diet.

Clinical and psychosocial outcomes, such as quality of life, were reported in the included studies. Significant effects on psychosocial outcomes were reported only for one intervention. However, positive clinical outcomes on activity levels, disease-specific self-care, and self-monitoring behaviors were apparent across all interventions. These findings present a similar view to this review and indicate that evidence-based digital interventions, often provided through web- or app-based delivery, have the potential to promote positive behavior change

XSL•FO
RenderX

and better support the self-management of conditions when delivered with correct guidance and tailored to the individual.

Conclusions

This review explored and examined evidence concerning web- and app-based interventions targeted at self-management of dementia through a narrative synthesis methodology. Many of the interventions reviewed had a positive impact on the self-management concept they were targeting, which suggests that their use could prove beneficial in dementia care. The successful adoption of these interventions appears to be dependent on individuals' engagement and their confidence in using the technology. Common factors influencing nonadoption appear to be a lack of confidence or familiarity with using technology, apprehension about encountering and resolving technological difficulties, and forgetting to use the intervention.

The findings are beneficial to health services and policy makers in considering how to incorporate self-management in dementia care and to researchers to help design better studies on the effectiveness of web- and app-based interventions. Barriers to adoption and implementation should be considered when delivering these interventions digitally to maximize the potential reach and effect on people living with dementia and their families. Conclusions drawn from this review will provide a positive contribution to the growing evidence base and increase the understanding of the use of these types of interventions in the self-management of dementia and their role in service provision.

Acknowledgments

This project was funded by the National Institute for Health Research Applied Research Collaboration East Midlands. The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Authors' Contributions

ARL conducted the review and narrative synthesis and wrote the manuscript. EVG provided a comprehensive second review during all stages of the review screening process. OM and MO provided significant feedback on the narrative synthesis process and the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Details of study characteristics.
[DOCX File , 16 KB - jmir_v23i7e26551_app1.docx]

Multimedia Appendix 2
Table S2. Details of study interventions.
[DOCX File , 16 KB - jmir_v23i7e26551_app2.docx]

Multimedia Appendix 3
Table S3. Outcome measures and key findings from each included study.
[DOCX File , 16 KB - jmir_v23i7e26551_app3.docx]

References

1. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774 [FREE Full text] [doi: 10.2196/16774] [Medline: 33112239]
2. Neal D, van den Berg F, Planting C, Ettema T, Dijkstra K, Finnema E, et al. Can use of digital technologies by people with dementia improve self-management and social participation? A systematic review of effect studies. J Clin Med 2021 Feb 5;10(4):604 [FREE Full text] [doi: 10.3390/jcm10040604] [Medline: 33562749]
3. Wittenberg R, Hu B, Barraza-Araiza L, Rehill A. Projections of Older People With Dementia and Costs of Dementia Care in the United Kingdom, 2019–2040. Care Policy and Evaluation Centre, London School of Economics and Political Sciences. 2019. URL: https://www.alzheimers.org.uk/sites/default/files/2019-11/cpec_report_november_2019.pdf [accessed 2020-06-23]
4. What is Dementia? Dementia UK. URL: https://www.dementiauk.org/understanding-dementia/what-is-dementia/ [accessed 2020-05-29]
5. How Many People Have Dementia and What Is the Cost of Dementia Care? Alzheimer's Society. 2019. URL: https://www.alzheimers.org.uk/about-us/policy-and-influencing/dementia-scale-impact-numbers [accessed 2020-06-30]

XSL•FO
RenderX

6. Whitlatch CJ, Orsulic-Jeras S. Meeting the informational, educational, and psychosocial support needs of persons living with dementia and their family caregivers. Gerontologist 2018 Jan 18;58(suppl_1):S58-S73. [doi: 10.1093/geront/gnx162] [Medline: 29361068]
7. Prime Minister's Challenge on Dementia 2020. Department of Health. 2015. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/414344/pm-dementia2020.pdf [accessed 2027-05-20]
8. Taylor S, Pinnock H, Epiphaniou E, Pearce G, Parke H, Schwappach A, et al. A Rapid Synthesis of the Evidence on Interventions Supporting Self-Management for People With Long-term Conditions: Prisms: Practical Systematic Review of Self-management Support for Long-term Conditions. Southampton, UK: NIHR Journals Library; 2014.
9. de Longh A, Fagan P, Fenner J, Kidd L. A Practical Guide to Self-Management Support: Key Components for Successful Implementation. London, England, UK: Health Foundation; 2015.
10. Supported Self-Management. NHS England. URL: https://www.england.nhs.uk/personalisedcare/supported-self-management/ [accessed 2020-02-18]
11. Hibbard JH, Gilburt H. Supporting People to Manage Their Health: An Introduction to Patient Activation. London, England, UK: The King's Fund; May 2014.
12. Oyebode J, Parveen S. Psychosocial interventions for people with dementia: an overview and commentary on recent developments. Dementia 2019 Jan;18(1):8-35. [doi: 10.1177/1471301216656096] [Medline: 27380931]
13. Olazarán J, Reisberg B, Clare L, Cruz I, Peña-Casanova J, del Ser T, et al. Nonpharmacological therapies in alzheimer's disease: a systematic review of efficacy. Dement Geriatr Cogn Disord 2010;30(2):161-178 [FREE Full text] [doi: 10.1159/000316119] [Medline: 20838046]
14. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. J Med Internet Res 2020 Sep 9;22(9):e19468 [FREE Full text] [doi: 10.2196/19468] [Medline: 32902388]
15. Wey S. One size does not fit all: person-centred approaches to the use of assistive technology. In: Marshall M, editor. Perspectives on Rehabilitation and Dementia. London, England, UK: Jessica Kingsley Publishers; 2004:202-210.
16. Lim FS, Wallace T, Luszcz MA, Reynolds KJ. Usability of tablet computers by people with early-stage dementia. Gerontology 2013;59(2):174-182. [doi: 10.1159/000343986] [Medline: 23257664]
17. het Veld JG, Verkaik R, van Meijel B, Francke AL. A systematic meta-review of self-management support for people with dementia. Dementia 2020 Feb;19(2):253-269 [FREE Full text] [doi: 10.1177/1471301218772894] [Medline: 29699409]
18. Rai H, Griffiths R, Yates L, Schneider J, Orrell M. Field-testing an iCST touch-screen application with people with dementia and carers: a mixed method study. Aging Ment Health 2021 Jun;25(6):1008-1018. [doi: 10.1080/13607863.2020.1783515] [Medline: 32578445]
19. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012 Feb 9;1:2 [FREE Full text] [doi: 10.1186/2046-4053-1-2] [Medline: 22587842]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
22. Reilly S, Miranda-Castillo C, Malouf R, Hoe J, Toot S, Challis D, et al. Case management approaches to home support for people with dementia. Cochrane Database Syst Rev 2015 Jan 5;1:CD008345 [FREE Full text] [doi: 10.1002/14651858.CD008345.pub2] [Medline: 25560977]
23. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York, England, UK: University of York NHS Centre for Reviews & Dissemination; Jan 2009.
24. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Commun Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
25. Trac MH, McArthur E, Jandoc R, Dixon SN, Nash DM, Hackam DG, et al. Macrolide antibiotics and the risk of ventricular arrhythmia in older adults. Can Med Assoc J 2016 Apr 19;188(7):E120-E129 [FREE Full text] [doi: 10.1503/cmaj.150901] [Medline: 26903359]
26. CASP (Qualitative) Checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2020-02-17]
27. O'Connor SR, Tully MA, Ryan B, Bradley JM, Baxter GD, McDonough SM. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. BMC Res Notes 2015 Jun 6;8:224 [FREE Full text] [doi: 10.1186/s13104-015-1181-1] [Medline: 26048813]
28. Stansfeld J, Crellin N, Orrell M, Wenborn J, Charlesworth G, Vernooij-Dassen M. Factors related to sense of competence in family caregivers of people living with dementia in the community: a narrative synthesis. Int Psychogeriatr 2019 Jun;31(6):799-813 [FREE Full text] [doi: 10.1017/S1041610218001394] [Medline: 30466499]
29. McDermott O, Crellin N, Ridder HM, Orrell M. Music therapy in dementia: a narrative synthesis systematic review. Int J Geriatr Psychiatry 2013 Aug;28(8):781-794. [doi: 10.1002/gps.3895] [Medline: 23080214]

XSL•FO
RenderX

30. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews A Product from the ESRC Methods Programme. Lancaster University. 2006 Apr. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2020-02-01]
31. Øksnebjerg L, Woods B, Ruth K, Lauridsen A, Kristiansen S, Holst HD, et al. A tablet app supporting self-management for people with dementia: explorative study of adoption and use patterns. JMIR Mhealth Uhealth 2020 Jan 17;8(1):e14694 [FREE Full text] [doi: 10.2196/14694] [Medline: 31951217]
32. Thorpe J, Forchhammer BH, Maier AM. Adapting mobile and wearable technology to provide support and monitoring in rehabilitation for dementia: feasibility case series. JMIR Form Res 2019 Oct 17;3(4):e12346 [FREE Full text] [doi: 10.2196/12346] [Medline: 31625951]
33. Kerssens C, Kumar R, Adams AE, Knott CC, Matalenas L, Sanford JA, et al. Personalized technology to support older adults with and without cognitive impairment living at home. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):85-97 [FREE Full text] [doi: 10.1177/1533317514568338] [Medline: 25614507]
34. Kerkhof Y, Pelgrum-Keurhorst M, Mangiaracina F, Bergsma A, Vrauwdeunt G, Graff M, et al. User-participatory development of FindMyApps; a tool to help people with mild dementia find supportive apps for self-management and meaningful activities. Digit Health 2019;5:2055207618822942 [FREE Full text] [doi: 10.1177/2055207618822942] [Medline: 30944726]
35. Perilli V, Lancioni GE, Singh NN, O'Reilly MF, Sigafoos J, Cassano G, et al. Persons with Alzheimer's disease make phone calls independently using a computer-aided telephone system. Res Dev Disabil 2012;33(4):1014-1020. [doi: 10.1016/j.ridd.2012.01.007] [Medline: 22502825]
36. Perilli V, Lancioni GE, Laporta D, Paparella A, Caffò AO, Singh NN, et al. A computer-aided telephone system to enable five persons with Alzheimer's disease to make phone calls independently. Res Dev Disabil 2013 Jun;34(6):1991-1997. [doi: 10.1016/j.ridd.2013.03.016] [Medline: 23584179]
37. Lancioni G, Singh N, O'Reilly M, Sigafoos J, D'Amico F, Pinto K, et al. A technology-aided program for helping persons with Alzheimer's disease perform daily activities. J Enabling Technol 2017 Sep 18;11(3):85-91. [doi: 10.1108/jet-03-2017-0011]
38. Lancioni GE, Singh NN, O'Reilly MF, Sigafoos J, D'Amico F, Laporta D, et al. Technology-based behavioral interventions for daily activities and supported ambulation in people with Alzheimer's disease. Am J Alzheimers Dis Other Demen 2018 Aug;33(5):318-326 [FREE Full text] [doi: 10.1177/1533317518775038] [Medline: 29742903]
39. Lancioni GE, Belardinelli MO, Singh NN, O'Reilly MF, Sigafoos J, Alberti G. Recent technology-aided programs to support adaptive responses, functional activities, and leisure and communication in people with significant disabilities. Front Neurol 2019;10:643 [FREE Full text] [doi: 10.3389/fneur.2019.00643] [Medline: 31312169]
40. McGoldrick C, Crawford S, Evans JJ. MindMate: a single case experimental design study of a reminder system for people with dementia. Neuropsychol Rehabil 2021 Jan;31(1):18-38. [doi: 10.1080/09602011.2019.1653936] [Medline: 31429370]
41. Boman I, Lundberg S, Starkhammar S, Nygård L. Exploring the usability of a videophone mock-up for persons with dementia and their significant others. BMC Geriatr 2014 Apr 16;14:49 [FREE Full text] [doi: 10.1186/1471-2318-14-49] [Medline: 24739662]

Abbreviations

CASP: Critical Appraisal Skills Program
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
RCT: randomized controlled trial
SMI: self-management intervention

Edited by R Kukafka; submitted 17.12.20; peer-reviewed by M Mulvenna, C Reidy; comments to author 28.02.21; revised version received 07.04.21; accepted 06.05.21; published 26.07.21.

Please cite as:
Lee AR, Gerritzen EV, McDermott O, Orrell M
Exploring the Role of Web-Based Interventions in the Self-management of Dementia: Systematic Review and Narrative Synthesis
J Med Internet Res 2021;23(7):e26551
URL: https://www.jmir.org/2021/7/e26551
doi:10.2196/26551
PMID:34309575

XSL•FO
RenderX

©Abigail Rebecca Lee, Esther Vera Gerritzen, Orii McDermott, Martin Orrell. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Effectiveness of eHealth Interventions in Improving Medication Adherence for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review

Mieke H J Schulte[1*], PhD; Jiska J Aardoom[1,2,3*], PhD; Lisa Loheide-Niesmann[1,4], MSc; Leonie L L Verstraete[1], MSc; Hans C Ossebaard[5], PhD; Heleen Riper[1,6,7], Prof Dr

[1]Department of Clinical, Neuro- and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands

[2]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[3]National eHealth Living Lab, Leiden, Netherlands

[4]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

[5]Dutch National Healthcare Institute, Diemen, Netherlands

[6]Department of Psychiatry, Amsterdam University Medical Center (VUmc), Vrije Universiteit, Amsterdam, Netherlands

[7]GGZ inGeest Specialized Mental Health Care, Research and Innovation, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Heleen Riper, Prof Dr
Department of Clinical, Neuro- and Developmental Psychology
Vrije Universiteit
Medical Faculty
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 20 59 82619
Email: h.riper@vu.nl

Abstract

Background: Poor treatment adherence in patients with chronic obstructive pulmonary disease (COPD) or asthma is a global public health concern with severe consequences in terms of patient health and societal costs. A potentially promising tool for addressing poor compliance is eHealth.

Objective: This review investigates the effects of eHealth interventions on medication adherence in patients with COPD or asthma.

Methods: A systematic literature search was conducted in the databases of Cochrane Library, PsycINFO, PubMed, and Embase for studies with publication dates between January 1, 2000, and October 29, 2020. We selected randomized controlled trials targeting adult patients with COPD or asthma, which evaluated the effectiveness of an eHealth intervention on medication adherence. The risk of bias in the included studies was examined using the Cochrane Collaboration's risk of bias tool. The results were narratively reviewed.

Results: In total, six studies focusing on COPD and seven focusing on asthma were analyzed. Interventions were mostly internet-based or telephone-based, and could entail telemonitoring of symptoms and medication adherence, education, counseling, consultations, and self-support modules. Control groups mostly comprised usual care conditions, whereas a small number of studies used a face-to-face intervention or waiting list as the control condition. For COPD, the majority of eHealth interventions were investigated as an add-on to usual care (5/6 studies), whereas for asthma the majority of interventions were investigated as a standalone intervention (5/7 studies). Regarding eHealth interventions targeting medication adherence for COPD, two studies reported nonsignificant effects, one study found a significant effect in comparison to usual care, and three reported mixed results. Of the seven studies that investigated eHealth interventions targeting medication adherence in asthma, three studies found significant effects, two reported nonsignificant effects, and two reported mixed effects.

Conclusions: The mixed results on the effectiveness of eHealth interventions in improving treatment adherence for asthma and COPD are presumably related to the type, context, and intensity of the interventions, as well as to differences in the

XSL•FO
RenderX

operationalization and measurement of adherence outcomes. Much remains to be learned about the potential of eHealth to optimize treatment adherence in COPD and asthma.

(J Med Internet Res 2021;23(7):e29475) doi:10.2196/29475

KEYWORDS

chronic obstructive pulmonary disease; asthma; medication adherence; exercise adherence; treatment adherence; eHealth; systematic review; COPD; adherence; exercise; treatment; review

Introduction

With a global prevalence of over 299 million people living with chronic obstructive pulmonary disease (COPD) and almost 273 million people living with asthma in 2017 [1], COPD and asthma are common chronic lung diseases. They are a worldwide public health concern and they increasingly affect the lives of patients due to climate change and pollution [2]. The clinical and economic burden of asthma and COPD have been widely established [3]. Both these respiratory diseases are typically treated and managed with drug therapies, often in the form of daily inhaled medication. Full adherence is important for optimal management and treatment of COPD and asthma [4-6]. This is especially the case when patients become more vulnerable, such as during environmental disruptions or the current COVID-19 pandemic [7].

Unfortunately, adherence to treatment regimens for COPD and asthma is often poor. Adherence to inhaled corticosteroids (ICS) in adult patients with asthma is estimated to range from 22% to 63% [8]. Medication adherence in patients with COPD reportedly ranges from 0.3% to 68%, depending on the type and combination of medications [9]. Adherence is a complex, multifaceted concept, including many potential contributing factors, which can be medication-related (eg, side effects) or patient-related (eg, forgetfulness, medication beliefs) [10]. Poor adherence can severely impact patients' health outcomes, with consequences including an increased risk of mortality and exacerbations, as well as diminished disease control and quality of life [5]. Poor adherence has furthermore been associated with higher health care utilization and costs [5,6]. Therefore, there is an urgent need for interventions that can improve treatment adherence in individuals with COPD or asthma.

A recent Cochrane review comprising 28 randomized controlled trials reported positive effects of various interventions to improve adherence to ICS in asthma in comparison to usual care [11]. After approximately 71 weeks of follow-up, a 20% improvement was achieved for people who were given education about adherence (20 trials) or who were provided with electronic monitoring or reminders to use their inhaler (11 trials). Another review investigated the effect of interventions to improve medication adherence in COPD [12]. Overall, five of the seven studies reported significant improvements in adherence. Effective strategies involved brief counseling, monitoring, and feedback on adherence through electronic medication delivery devices, as well as multicomponent interventions including education, self-management, motivational interviewing, and extra support (eg, clinic visits, phone calls) by health care professionals. Whether such strategies produced effects of similar magnitudes remains unclear.

Increasingly, eHealth is being used in the provision of health care services such as patient communication, monitoring, and education. In general, eHealth can be an effective tool to address poor treatment adherence in patients with chronic diseases, as indicated by the results of numerous studies focusing on different target populations [13-17]. However, to our knowledge, no systematic review or meta-analysis has yet been performed investigating the effects of eHealth interventions on adherence specifically for patients with COPD and only limited research has focused on youth or adult asthma populations. Bonini [18] conducted a systematic review of the literature published in 2016 that assessed the effects of eHealth on asthma management, incorporating multiple components including medication adherence in children and adults. The findings suggested an overall beneficial effect of eHealth on asthma control and management, whereby eHealth included mobile health systems (mHealth), telemedicine, electronic health records, and digital app interventions. A recent meta-analysis investigated the effect of eHealth on ICS adherence in patients with asthma (including both children and adults), comprising 15 randomized controlled trials with a total of 13,907 participants. Compared with usual care, a small but significant overall effect of eHealth interventions was observed. In addition, a pooled analysis of four studies provided evidence for the superiority of mHealth interventions such as SMS text messages and audiovisual reminders as compared to usual care [19].

The aim of our study was to systematically review the effectiveness of eHealth interventions in improving medication adherence in adult patients with COPD or asthma. In this review, the results were presented separately for asthma and COPD in order to enable investigation of potential differences between the diseases, as well as to allow for potential nuance in terms of the effectiveness of specific eHealth interventions for the two diseases separately.

Methods

No review protocol was made beforehand and the review was not registered in any database or registry. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist can be found in Multimedia Appendix 1.

Search Strategy

Our search strategy was part of a broader search performed in a research project on the role of eHealth in treatment adherence in chronic lung disease, including obstructive sleep apnea (OSA), asthma, and COPD. The results regarding OSA have been published elsewhere [20]. The search was conducted in the electronic databases of the Cochrane Library (Wiley),

XSL•FO
RenderX

PsycINFO (EBSCO), PubMed, and Embase. The search results were limited to available full-text articles in English or Dutch with publication dates from January 1, 2000, to October 29, 2020. Numerous terms related to eHealth technology, patient adherence, and the target populations (asthma, COPD) were combined, using both free-text and index terms (for the full search string, see Multimedia Appendix 2). In addition, reference lists of the included studies, as well as relevant systematic reviews, were checked for potentially relevant additional studies.

Study Selection and Data Extraction

Inclusion criteria were as follows: (1) The target population comprised patients aged ≥18 years with COPD or asthma. (2) One or more main component(s) of the intervention were delivered by eHealth technology, or an eHealth component was investigated as an add-on intervention to usual care. The criteria to qualify as an eHealth intervention were that (A) the intervention was delivered via information and communications technology such as telephone calls, telemedicine (eg, videoconferencing), websites, smartphone apps, or SMS text messages; and (B) the intervention was delivered independently of time and place (eg, videos delivered in face-to-face sessions were not considered eligible). (3) Intervention effects were compared to a control group, with exclusion of control conditions containing the same eHealth component as the experimental condition. (4) Outcomes were assessed in terms of at least one quantitative measure of adherence to the medical treatment—that is, to oral or inhaler medication. (5) Adherence measures were compared statistically between study conditions. (6) The study design was a randomized controlled trial.

All titles and abstracts were independently screened by two reviewers (JA and LL: January 1, 2000, through March 20, 2018; MS and LV: March 21, 2018, through October 29, 2020). Subsequently, the full-text articles of the selected papers were screened to determine eligibility for this review. Covidence software was used to manage the screening process and the risk-of-bias assessments. Data on study reference, design, population, interventions, outcomes, and results were extracted by means of a data extraction form in an Excel spreadsheet by JA (January 1, 2000, through March 20, 2018) and MS (March 21, 2018, through October 29, 2020).

Quality Assessment

The Cochrane Collaboration's risk-of-bias tool [21] was used to assess the quality of all included studies. Two reviewers (JA and LL or MS and LV, depending on publication date; see above) independently evaluated the following dimensions of risk of bias: (1) adequacy of random sequence generation, (2) adequacy of concealment of allocation sequence to personnel, (3) blinding of study participants and personnel, (4) blinding of outcome assessors, (5) adequacy of handling of incomplete outcome data, and (6) selective outcome reporting. Each study was rated per dimension as "low risk," "high risk," or "unclear risk." Disagreements between reviewers were resolved by discussion.

Data Analysis

Due to the limited number of available studies and the heterogeneity of the studies in terms of designs and characteristics, as well as the assessments and operationalizations of medication adherence, the results were narratively reviewed and no meta-analysis has been performed.

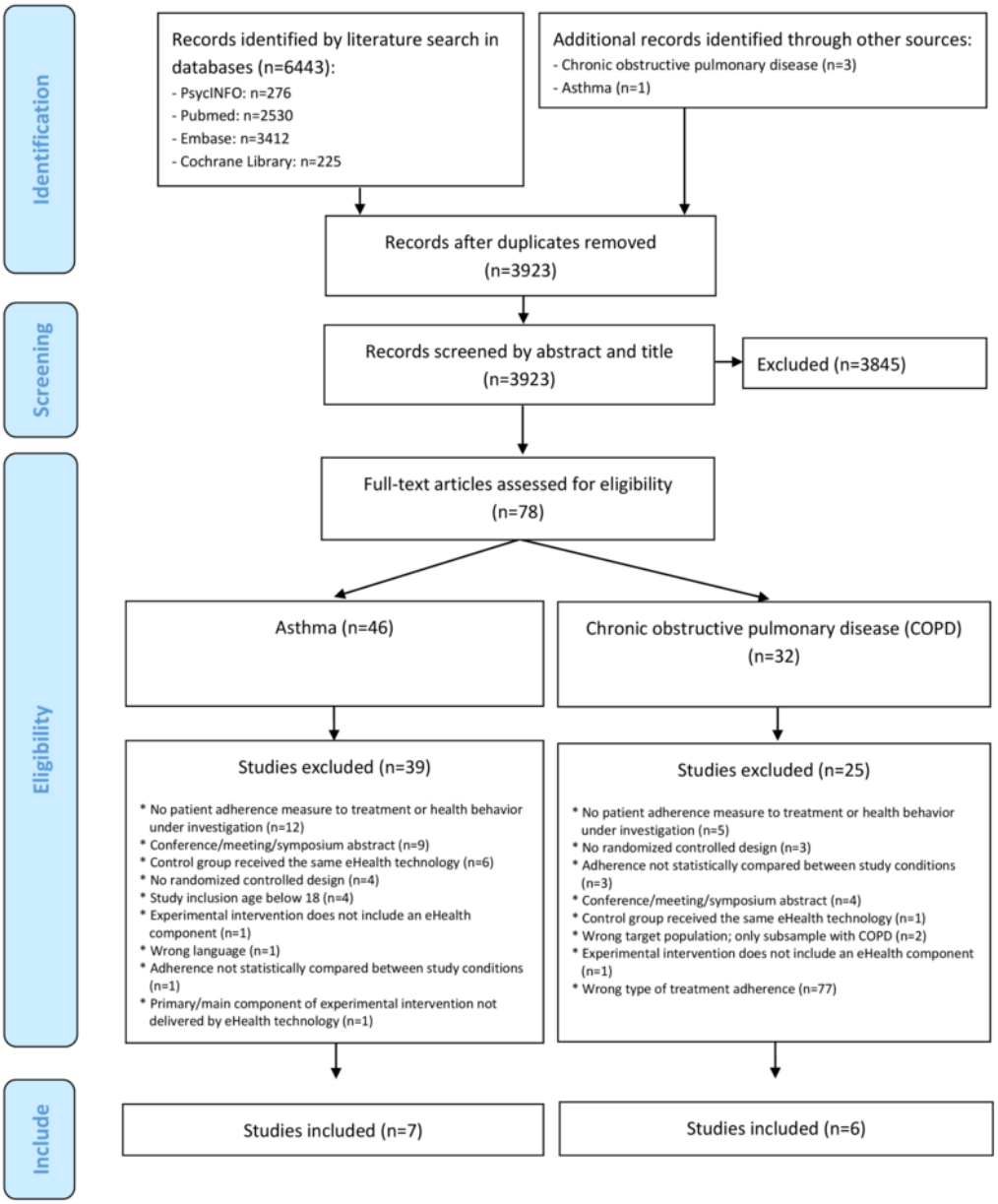
Results

Search and Screening

Figure 1 depicts the PRISMA flow diagram of study identification and selection. The pooled systematic search resulted in a total of 6447 potentially relevant articles covering COPD or asthma. After removal of 2520 duplicates, a total of 3923 articles were selected for title and abstract screening. A total of 78 studies were then selected for full-text screening. Of these, 32 targeted COPD and 46 targeted asthma. Full-text screening of the eligibility criteria eventually led to the inclusion of 6 studies targeting COPD and 7 studies targeting asthma.

XSL•FO
RenderX

Figure 1. PRISMA flowchart describing study identification and selection process. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Results for COPD

Study Characteristics

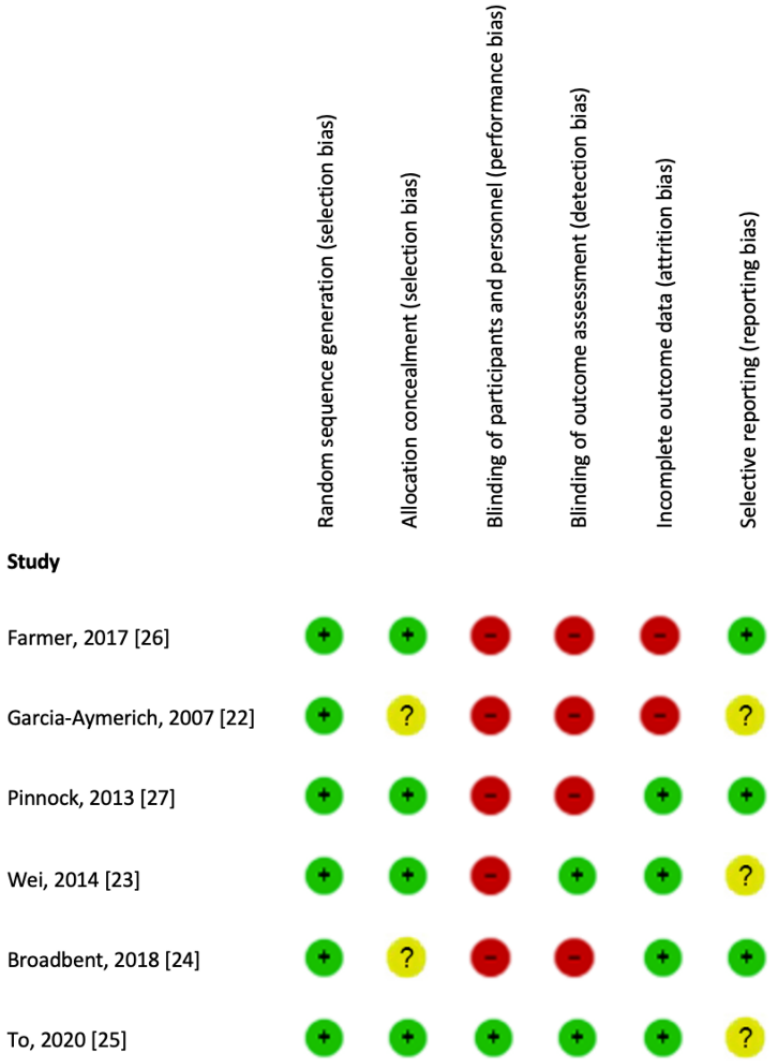
Multimedia Appendix 3 provides an overview of study and intervention characteristics; of the six included studies, two interventions were internet-based and four were telephone-based. Interventions involved the telemonitoring of symptoms and adherence, education and counseling (eg, knowledge of the disease, smoking cessation, inhaler techniques), and self-support modules (eg, to help patients identify disease exacerbations or to support psychological well-being). Medication adherence was a primary outcome in three studies and a secondary outcome in three others. The studies were conducted in Europe, China, and New Zealand.

Quality Assessment

Figure 2 presents the results of the risk-of-bias assessment for each study separately. None of the studies were rated as having low risk of bias on all six dimensions. The majority of studies had a high risk of performance bias (n=5) and detection bias (n=4). These were due to the lack of blinding and participants self-reporting their medication adherence while being aware of their allocated study condition. Studies with a high risk of attrition bias (n=2) generally did not analyze the data according to an intention-to-treat design, thus excluding participants who did not adhere to the intervention or were lost to follow-up.

XSL•FO
RenderX

Figure 2. Risk of bias analysis on individual studies investigating effectiveness of eHealth interventions on medication adherence in patients with chronic obstructive pulmonary disease.



Effects of eHealth Interventions on Medication Adherence

The effectiveness of the interventions is shown in Table 1. Of the six included studies, four studies focusing on medication adherence in COPD demonstrated the eHealth intervention to be more effective than care as usual in terms of improving medication adherence on at least one of the operationalizations of adherence [22-25]. However, three of these four studies reported nonsignificant effects when other operationalizations of medication adherence were used [22,24,25]. In addition, two studies found no differences between participants who received the eHealth intervention (ie, online monitoring and as-needed telephone contact) supplemental to care as usual, or care as usual only [26,27]. Both of these studies used the self-report Medication Adherence Rating Scale (MARS).

Telephone-based pharmaceutical care including education and counseling was found to be effective as compared to care as usual in terms of adherence operationalized as pill count [23]. The three studies that reported mixed results all investigated the effect of an eHealth intervention as an add-on to care as usual. These studies differed regarding the type of intervention and the technology used, as well as how medication adherence was operationalized. In one study, telephone-based integrated care resulted in significant effects in terms of the percentages of self-reported medication inhaler adherence and observed correct inhaler maneuvers, whereas no effects were found for the percentages of self-reported oral medication adherence [22]. In another study, telephone-based telemonitoring did not result in significant effects on adherence to the medication regimen, but there were significant effects on the percentages of people adhering at least 80% to the regimen; in both those operationalizations, adherence was measured objectively by an administration tracker attached to the device [25]. In the last study, telemonitoring and treatment reminders delivered by an internet-linked robot resulted in significant results when medication adherence was measured with a self-reported questionnaire, but nonsignificant results when the percentages of medication adherence were measured objectively by an administration tracker attached to the device [24].

Results for Asthma

Study Characteristics

Multimedia Appendix 3 presents the details on study and intervention characteristics. Of the seven included studies focusing on asthma, six interventions were telephone-based

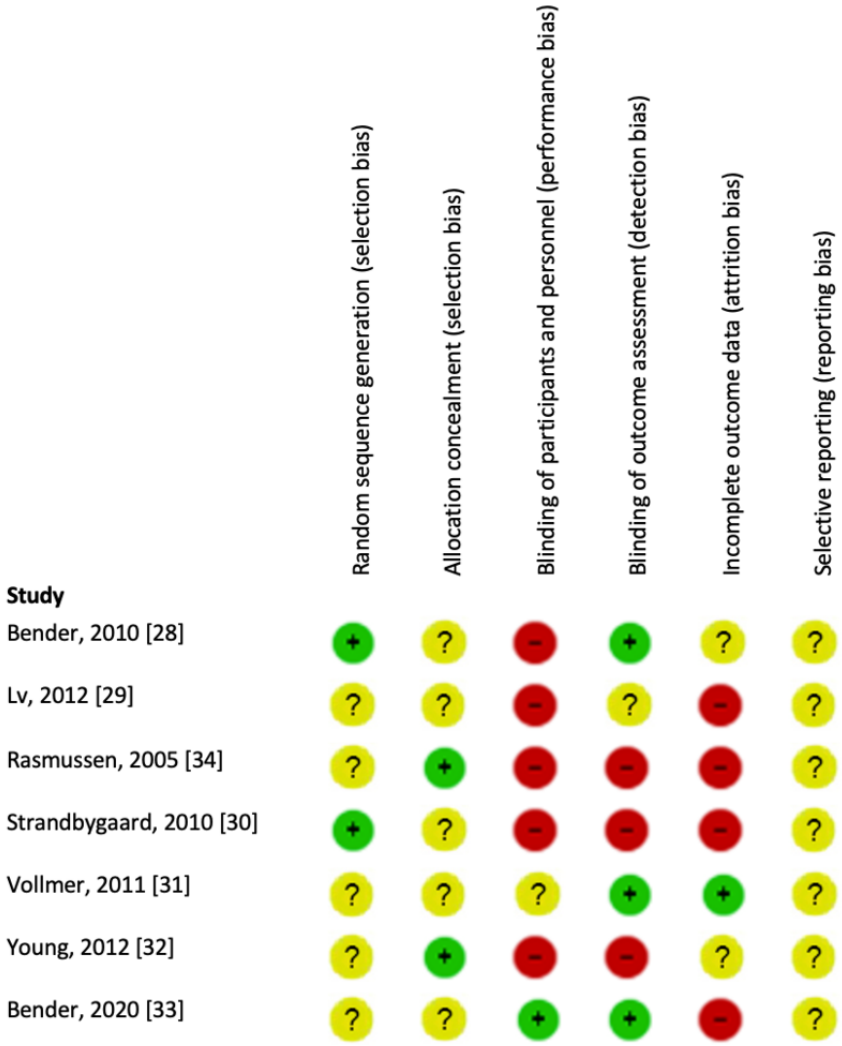
XSL•FO
RenderX

[28-33] and one was internet-based [34]. Interventions involved monitoring and management of medication, including reminders for intake and refills, as well as pharmacist consultations. Control conditions included, among others, asthma education, monitoring and treatment by a general practitioner or specialist, and inhaler use tracking. Study periods ranged from 2.5 to 18 months. Overall, five studies included medication adherence as a primary outcome. Most studies were conducted in the United States (n=4).

Quality Assessment

Figure 3 presents the results of the risk-of-bias assessment. None of the studies were rated as having low risk of bias on all six dimensions. A substantial number of studies had a high risk of performance bias (n=5) or detection bias (n=3), due to the lack of blinding. Most studies also had a high risk of attrition bias (n=4), mainly because they did not analyze the data in an intention-to-treat design.

Figure 3. Risk of bias analysis on individual studies investigating effectiveness of eHealth interventions on medication adherence in patients with asthma.



Effects of eHealth Interventions on Medication Adherence

The effectiveness of the interventions is displayed in Table 1. We determined that five of the seven included studies on medication adherence in asthma reported significant effects on at least one of the operationalizations of medication adherence compared to control conditions [28,30,31,33,34]. However, at the same time, two of these five studies reported nonsignificant effects when other operationalizations of medication adherence were used [30,33]. In addition, two studies reported nonsignificant effects [29,32]. In two studies, a standalone eHealth intervention involving an interactive voice response system for monitoring, medication reminders, and education was found to significantly increase self-reported medication adherence [28,31]. One of these studies operationalized medication adherence on the basis of self-report [31], whereas the other tracked inhaler use or assessed inhaler weight [28]. In another study, a guided standalone internet-based monitoring and management tool was found to significantly increase self-reported medication adherence as compared to a control group receiving monitoring and treatment by a general practitioner or specialist [34]. Another study also investigating the effect of SMS text message medication reminders as an add-on to care as usual reported mixed results depending on how adherence was operationalized [30]. Compared to care as

XSL•FO
RenderX

usual, there was a significant improvement in the percentage of inhaler use, but not in the percentage of participants collecting their medication refills, as recorded in pharmacy reports. Another study also investigated the effect of a standalone monitoring and management support tool, but provided via SMS text message or telephone interactive voice response or via email [33]; the results were mixed depending on the operationalization of medication adherence and the group comparisons. Comparing the combined intervention groups with care as usual, the study found no significant effects on self-reported medication adherence. When the intervention groups were compared, no significant effect was found on self-reported use of reliever medication or on the asthma medication ratio, but a greater increase in self-reported use of controller medication was found in the SMS text message/telephone group as compared to the email group. Telephone-based pharmacist consultations as an add-on to care as usual did not result in a significant increase in self-reported adherence rates compared with care as usual [32]. Lv and colleagues [29] investigated the effect of a standalone eHealth intervention using SMS reminders about asthma management; they reported nonsignificant effects in comparison with care as usual or with verbal or written asthma education. Medication adherence was defined as the percentage of participants that were adherent, although the exact operationalization was unclear.

XSL•FO
RenderX

Table 1. Effects of eHealth interventions for medication adherence in chronic obstructive pulmonary disease and asthma: study results[a].

Condition; reference and study design; and outcome operationalization	Outcome measure	Between-group results	Between-group statistic (*P* value)	Study quality (number of dimensions unclear/low/high risk)[b]
Chronic obstructive pulmonary disease				
Farmer, 2017 [26], add-on to care as usual			.77	0/3/3
Medication adherence	Medication Adherence Rate Scale	Difference in score from baseline to 12 months. TG: 0.17 (2.47) vs CG: 0.33 (3.65).		
Garcia-Aymerich, 2007 [22], add-on to care as usual				2/1/3
Percentage of oral medication adherence	Medication Adherence Scale	12-month results. TG: 90% vs CG: 85%	.57	
Percentage of inhaler medication adherence	Inhaler Adherence Scale	12-month results. TG: 71% vs CG: 37%	.009	
Percentage of correct inhaler maneuvers	Observed inhaler skills	12-month results. TG: 86% vs CG: 24%	.001	
Pinnock, 2013 [27], add-on to care as usual				0/4/2
Medication adherence	Medication Adherence Rate Scale	12-month results. TG: 24.0 (1.7) vs CG: 23.7 (1.9)	.05	
Wei, 2014 [23], standalone				1/4/1
Percentage of medication adherence	Pill count	12-month results. TG: 66.5 (8.6) vs CG: 54.4 (12.5)	.04	
Broadbent, 2018 [24], add-on to care as usual				1/3/2
Percentage of medication adherence	Administration tracker on device	TG: 48.5% vs CG: 29.5%	.03	
Medication adherence	Medication Adherence Rate Scale	Difference in score from baseline to 4 months. TG: 1.63 (0.56) vs CG: 0.12 (0.55)	.06	
To, 2020 [25], add-on to care as usual				1/5/0
Medication regimen adherence	Ratio number of doses taken (administration tracker) / number of doses prescribed	2-month results. TG: 99.8 (15.0) vs CG: 92.7 (30.0)	.12	
Percentage of ≥80% adherence	Ratio number of doses taken (administration tracker)/number of doses prescribed	2-month results. TG: 85.7 vs CG: 71.4	.02	
Correct inhaler technique	Inhaler use checklist	2-month results. TG: 91.9 (7.8) vs CG: 79.9 (17.1)	.002	
Asthma				
Bender, 2010 [28], standalone				3/2/1
Percentage of inhaled corticosteroids adherence (number of taken puff/number of prescribed puffs)	Electronic tracking device	CG: 64.5 (17.2) vs TG: 49.1 (16.8)	.003	
Lv, 2012 [29], add-on to care as usual				4/0/2
Percentage of adherers	Not specified	TG: 80 vs CG1: 74.1 vs CG2: 50	.11	
Rasmussen, 2005 [34], standalone				2/1/3
Percentage of inhaled corticosteroids adherers	Self-reported as almost always taking inhaled corticosteroids	TG: 87 vs CG2: 54	<.001	
Percentage of inhaled corticosteroids adherers	Self-reported as almost always taking inhaled corticosteroids	CG1: 79 vs CG2: 54	<.001	

XSL•FO
RenderX

Condition; reference and study design; and outcome operationalization	Outcome measure	Between-group results	Between-group statistic (*P* value)	Study quality (number of dimensions unclear/low/high risk)[b]
Strandbygaard, 2010 [30], add-on to care as usual				2/1/3
Percentage of inhaled corticosteroids adherence	Administration tracker on device	TG: 81.5 vs CG: 70.1	.02	
Percentage collecting medication	Pharmacy reports	TG: 64.3 vs CG: 66.7	.69	
Number of days until collecting medication	Pharmacy reports	TG: 32 (13-50) vs CG: 29 (13-49)[c]	.56	
Vollmer, 2011 [31], standalone			.002	4/2/0
Change in medication adherence	Difference in modified medication possession ratio	TG: 0.40 (0.32) vs CG: –0.04 (0.24)		
Young, 2012 [32], standalone			.07	3/1/2
Percentage of low adherers	Morisky Medication Adherence Scale	TG: 47.0 vs CG: 26.0		
Bender, 2020 [33], standalone				3/2/1
Reliever medication use	Number of beta-agonist canisters dispensed	6- to 18-month results. CG: 4.71 (0.23) vs TG1+TG2: 4.96 (0.16)	.19	
Reliever medication use	Number of beta-agonist canisters dispensed	6- to 18-month results: TG1: 5.15 (0.27) vs TG2: 4.76 (0.19)	.72	
Controller medication use	Number of inhaled corticosteroids canisters dispensed	6- to 18-month results: CG: 6.43 (0.26) vs TG1+TG2: 6.70 (0.20)	.19	
Controller medication use	Number of inhaled corticosteroids canisters dispensed	6- to 18-month results: TG1: 6.85 (0.31) vs TG2: 6.55 (0.24)	.03	
Ratio used medication	Asthma medication ratio	6- to 18-month results: CG: 0.58 (0.01) vs TG1+TG2: 0.57 (0.01)	.99	
Ratio used medication	Asthma medication ratio	6- to 18-month results. TG1: 0.57 (0.01) vs TG2: 0.58 (0.01)	.05	

[a]CG: control group; TG: treatment group.

[b]Risk of bias according to six dimensions of the Cochrane Collaboration's risk of bias tool.

[c]Values represent means (ranges).

Discussion

General Discussion

This review investigated the effects of eHealth interventions in improving adherence to medication treatment by patients with COPD or asthma. In general, mixed results were found and no definite conclusions could be drawn.

The mixed results of the current review may be explained by differences in study design, type and intensity of eHealth interventions, type of control condition, and assessment and operationalization of outcome measures. Previous reviews and meta-analyses have demonstrated small but significant effects of eHealth interventions in improving medication adherence for patients with asthma [18,19], as well as for more diverse populations of chronically ill patients [13] and individuals with long-term medication [35]. The latter two studies found positive effects in 66% and 59% of the studies, respectively. In line with our rather inconclusive and mixed findings, all of the abovementioned overview studies highlighted the considerable amount of heterogeneity among study designs and outcomes, and the limited number of high-quality studies conducted.

Differences in operationalization may have contributed to the discrepant results, in that self-report questionnaires such as the MARS [36] might be less sensitive in detecting actual changes in treatment adherence than more direct adherence assessments such as pill counts or inhaler use tracking devices. Indeed, such variations in objective and subjective measurements of medication adherence have been widely reported in the literature [37,38]. To the best of our knowledge, this is the first systematic review to investigate the effects of eHealth interventions on medication adherence in adult patients with COPD or asthma. Our findings are limited by the small number of included studies and considerable heterogeneity regarding different study aspects. This challenges the interpretation of eHealth intervention effects

XSL•FO
RenderX

in terms of medication adherence for patients with asthma and COPD.

Future Research Directions and Recommendations

Given the limited number of high-quality studies, more studies that minimize potential bias risks are needed to create a more substantial and reliable body of research on the effectiveness of eHealth interventions to improve treatment adherence in COPD and asthma. Furthermore, as a wide variety of outcome measures have been used, future studies could benefit from standardizing measures with respect to adherence outcome. In addition, standardizing the operationalizations of such outcome measures and reporting effect sizes instead of mere statistics in terms of significance could potentially lead to more clear-cut results. Future research would benefit from studies with sufficient statistical power. This would also allow for subgroup analyses, which could provide more insight into what types, intensities, and components of interventions might be more effective for different subgroups, ultimately leading to more personalized or tailor-made treatments. Preliminary research suggests that increased adherence can improve patient outcomes as well as reduce health care costs [5,39]. However, more research is needed to elucidate the cost- effectiveness of eHealth interventions targeting adherence in comparison to usual care. Finally, future studies of eHealth interventions should therefore incorporate cost-effectiveness analyses to elucidate effects in relation to costs as compared with usual care.

Conclusion

No firm conclusion can be drawn due to the small numbers of studies and their heterogeneous results. Much remains to be learned about the potential of eHealth in optimizing treatment adherence in COPD and asthma—for example, in terms of what types and intensities of eHealth intervention components are effective for what types of individuals.

Acknowledgments

This study was funded by the Dutch National Health Care Institute. The funder was not involved in carrying out the research activities. We would like to thank Caroline Planting for her help in conducting the literature search.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[DOCX File , 32 KB - jmir_v23i7e29475_app1.docx]

Multimedia Appendix 2
Search string.
[DOCX File , 21 KB - jmir_v23i7e29475_app2.docx]

Multimedia Appendix 3
Study and intervention characteristics.
[DOCX File , 20 KB - jmir_v23i7e29475_app3.docx]

References

1. GBD 2017 Disease and Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Dec 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Shin S, Bai L, Burnett RT, Kwong JC, Hystad P, van Donkelaar A, et al. Air Pollution as a Risk Factor for Incident Chronic Obstructive Pulmonary Disease and Asthma. A 15-Year Population-based Cohort Study. Am J Respir Crit Care Med 2021 May 01;203(9):1138-1148. [doi: 10.1164/rccm.201909-1744oc]
3. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med 2017 Sep;5(9):691-706 [FREE Full text] [doi: 10.1016/S2213-2600(17)30293-X] [Medline: 28822787]
4. Vrijens B, Dima AL, Van Ganse E, van Boven JF, Eakin MN, Foster JM, et al. What We Mean When We Talk About Adherence in Respiratory Medicine. J Allergy Clin Immunol Pract 2016 Sep;4(5):802-812 [FREE Full text] [doi: 10.1016/j.jaip.2016.05.019] [Medline: 27587314]
5. Mäkelä MJ, Backer V, Hedegaard M, Larsson K. Adherence to inhaled therapies, health outcomes and costs in patients with asthma and COPD. Respir Med 2013 Oct;107(10):1481-1490 [FREE Full text] [doi: 10.1016/j.rmed.2013.04.005] [Medline: 23643487]

XSL•FO
RenderX

6. van Boven JFM, Chavannes NH, van der Molen T, Rutten-van Mölken MPMH, Postma MJ, Vegter S. Clinical and economic impact of non-adherence in COPD: a systematic review. Respir Med 2014 Jan;108(1):103-113 [FREE Full text] [doi: 10.1016/j.rmed.2013.08.044] [Medline: 24070566]
7. Kaye L, Theye B, Smeenk I, Gondalia R, Barrett MA, Stempel DA. Changes in medication adherence among patients with asthma and COPD during the COVID-19 pandemic. J Allergy Clin Immunol Pract 2020 Jul;8(7):2384-2385 [FREE Full text] [doi: 10.1016/j.jaip.2020.04.053] [Medline: 32371047]
8. Bårnes CB, Ulrik CS. Asthma and adherence to inhaled corticosteroids: Current status and future perspectives. Respir Care 2015 Mar 12;60(3):455-468 [FREE Full text] [doi: 10.4187/respcare.03200] [Medline: 25118311]
9. Ingebrigtsen TS, Marott JL, Nordestgaard BG, Lange P, Hallas J, Dahl M, et al. Low use and adherence to maintenance medication in chronic obstructive pulmonary disease in the general population. J Gen Intern Med 2015 Jan 23;30(1):51-59 [FREE Full text] [doi: 10.1007/s11606-014-3029-0] [Medline: 25245885]
10. George M, Bender B. New insights to improve treatment adherence in asthma and COPD. PPA 2019 Jul;Volume 13:1325-1334. [doi: 10.2147/ppa.s209532]
11. Normansell R, Kew K, Stovold E. Interventions to improve adherence to inhaled steroids for asthma. Cochrane Database Syst Rev 2017 Apr 18;4(4):CD012226 [FREE Full text] [doi: 10.1002/14651858.CD012226.pub2] [Medline: 28417456]
12. Bryant J, McDonald V, Boyes A, Sanson-Fisher R, Paul C, Melville J. Improving medication adherence in chronic obstructive pulmonary disease: a systematic review. Respir Res 2013 Oct 20;14(1):109 [FREE Full text] [doi: 10.1186/1465-9921-14-109] [Medline: 24138097]
13. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
14. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
15. Tao D, Xie L, Wang T, Wang T. A meta-analysis of the use of electronic reminders for patient adherence to medication in chronic disease care. J Telemed Telecare 2014 Aug 21;21(1):3-13. [doi: 10.1177/1357633x14541041]
16. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile Telephone Text Messaging for Medication Adherence in Chronic Disease: A Meta-analysis. JAMA Intern Med 2016 Mar 1;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
17. Kini V, Ho PM. Interventions to Improve Medication Adherence: A Review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
18. Bonini M. Electronic health (e-Health): emerging role in asthma. Curr Opin Pulm Med 2017 Jan;23(1):21-26. [doi: 10.1097/MCP.0000000000000336] [Medline: 27763999]
19. Jeminiwa R, Hohmann L, Qian J, Garza K, Hansen R, Fox BI. Impact of eHealth on medication adherence among patients with asthma: A systematic review and meta-analysis. Respir Med 2019 Mar;149:59-68 [FREE Full text] [doi: 10.1016/j.rmed.2019.02.011] [Medline: 30803887]
20. Aardoom JJ, Loheide-Niesmann L, Ossebaard HC, Riper H. Effectiveness of eHealth Interventions in Improving Treatment Adherence for Adults With Obstructive Sleep Apnea: Meta-Analytic Review. J Med Internet Res 2020 Feb 18;22(2):e16972 [FREE Full text] [doi: 10.2196/16972] [Medline: 32130137]
21. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0). 2011. URL: https://handbook-5-1.cochrane.org/ [accessed 2021-07-16]
22. Garcia-Aymerich J, Hernandez C, Alonso A, Casas A, Rodriguez-Roisin R, Anto JM, et al. Effects of an integrated care intervention on risk factors of COPD readmission. Respir Med 2007 Jul;101(7):1462-1469 [FREE Full text] [doi: 10.1016/j.rmed.2007.01.012] [Medline: 17339106]
23. Wei L, Yang X, Li J, Liu L, Luo H, Zheng Z, et al. Effect of pharmaceutical care on medication adherence and hospital admission in patients with chronic obstructive pulmonary disease (COPD): a randomized controlled study. J Thorac Dis 2014 Jun;6(6):656-662 [FREE Full text] [doi: 10.3978/j.issn.2072-1439.2014.06.20] [Medline: 24976987]
24. Broadbent E, Garrett J, Jepsen N, Li Ogilvie V, Ahn HS, Robinson H, et al. Using Robots at Home to Support Patients With Chronic Obstructive Pulmonary Disease: Pilot Randomized Controlled Trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
25. To KW, Lee IF, Choi KC, Cheung YTY, Yu DSF. An information-motivation-behavioural-based model and adherence to inhalation therapy and other health outcomes in patients with chronic obstructive pulmonary disease: A pilot randomized controlled trial. Int J Nurs Pract 2020 Apr;26(2):e12799. [doi: 10.1111/ijn.12799] [Medline: 31903654]
26. Farmer A, Williams V, Velardo C, Shah SA, Yu L, Rutter H, et al. Self-Management Support Using a Digital Health System Compared With Usual Care for Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial. J Med Internet Res 2017 May 03;19(5):e144 [FREE Full text] [doi: 10.2196/jmir.7116] [Medline: 28468749]
27. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind,

XSL•FO
RenderX

multicentre, randomised controlled trial. BMJ 2013 Oct 17;347:f6070 [FREE Full text] [doi: 10.1136/bmj.f6070] [Medline: 24136634]
28. Bender BG, Apter A, Bogen DK, Dickinson P, Fisher L, Wamboldt FS, et al. Test of an interactive voice response intervention to improve adherence to controller medications in adults with asthma. J Am Board Fam Med 2010 Mar 05;23(2):159-165 [FREE Full text] [doi: 10.3122/jabfm.2010.02.090112] [Medline: 20207925]
29. Lv Y, Zhao H, Liang Z, Dong H, Liu L, Zhang D, et al. A mobile phone short message service improves perceived control of asthma: a randomized controlled trial. Telemed J E Health 2012 Jul;18(6):420-426. [doi: 10.1089/tmj.2011.0218] [Medline: 22667695]
30. Strandbygaard U, Thomsen SF, Backer V. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. Respir Med 2010 Feb;104(2):166-171 [FREE Full text] [doi: 10.1016/j.rmed.2009.10.003] [Medline: 19854632]
31. Vollmer W, Feldstein A, Smith D, Dubanoski J, Waterbury A, Schneider J, et al. Use of health information technology to improve medication adherence. Am J Manag Care 2011 Dec;17(12 Spec No):SP79-SP87 [FREE Full text] [Medline: 22216772]
32. Young HN, Havican SN, Griesbach S, Thorpe JM, Chewning BA, Sorkness CA. Patient and phaRmacist telephonic encounters (PARTE) in an underserved rural patient population with asthma: results of a pilot study. Telemed J E Health 2012;18(6):427-433 [FREE Full text] [doi: 10.1089/tmj.2011.0194] [Medline: 22656403]
33. Bender BG, Wagner NM, Shoup JA, Goodrich GK, Shetterly SM, Cvietusa PJ, et al. Adults With Asthma Experience No Increase in Asthma-related Exacerbations When Digital Communication Technology Tools Are Employed to Offset Provider Workload: A Pragmatic Randomized Trial. Med Care 2020;58(4):352-359. [doi: 10.1097/MLR.0000000000001265] [Medline: 32197029]
34. Rasmussen LM, Phanareth K, Nolte H, Backer V. Internet-based monitoring of asthma: a long-term, randomized clinical study of 300 asthmatic subjects. J Allergy Clin Immunol 2005 Jun;115(6):1137-1142. [doi: 10.1016/j.jaci.2005.03.030] [Medline: 15940125]
35. Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, et al. Effect of Interactive eHealth Interventions on Improving Medication Adherence in Adults With Long-Term Medication: Systematic Review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
36. Thompson K, Kulkarni J, Sergejew A. Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses. Schizophrenia Research 2000 May;42(3):241-247. [doi: 10.1016/s0920-9964(99)00130-9]
37. Van Steenis MNA, Driesenaar JA, Bensing JM, Van Hulten R, Souverein PC, Van Dijk L, et al. Relationship between medication beliefs, self-reported and refill adherence, and symptoms in patients with asthma using inhaled corticosteroids. Patient Prefer Adherence 2014 Jan;8:83-91 [FREE Full text] [doi: 10.2147/PPA.S44185] [Medline: 24470757]
38. Lam WY, Fresco P. Medication Adherence Measures: An Overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]
39. Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of medication adherence on hospitalization risk and healthcare cost. Med Care 2005 Jun;43(6):521-530. [doi: 10.1097/01.mlr.0000163641.86870.af] [Medline: 15908846]

Abbreviations

COPD: chronic obstructive pulmonary disease
ICS: inhaled corticosteroids
MARS: Medication Adherence Rating Scale
OSA: obstructive sleep apnea
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 08.04.21; peer-reviewed by B Pouls, I Yang; comments to author 27.04.21; revised version received 12.06.21; accepted 14.06.21; published 27.07.21.

Please cite as:
Schulte MHJ, Aardoom JJ, Loheide-Niesmann L, Verstraete LLL, Ossebaard HC, Riper H
Effectiveness of eHealth Interventions in Improving Medication Adherence for Patients With Chronic Obstructive Pulmonary Disease or Asthma: Systematic Review
J Med Internet Res 2021;23(7):e29475
URL: https://www.jmir.org/2021/7/e29475
doi:10.2196/29475
PMID:34313593

XSL•FO
RenderX

©Mieke H J Schulte, Jiska J Aardoom, Lisa Loheide-Niesmann, Leonie L L Verstraete, Hans C Ossebaard, Heleen Riper. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Use of Twitter by Medical Journals: Systematic Review of the Literature

Natalie Erskine[1], BSc, MSc; Sharief Hendricks[1,2,3], BSc, MSc, PhD

[1]Division of Physiological Sciences, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[2]Health, Physical Activity, Lifestyle, and Sport Research Centre, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[3]Carnegie Applied Rugby Research Centre, Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, United Kingdom

Corresponding Author:
Natalie Erskine, BSc, MSc
Division of Physiological Sciences
Department of Human Biology
Faculty of Health Sciences, University of Cape Town
Boundary Road
Newlands
Cape Town, 7725
South Africa
Phone: 27 0728716033
Email: natserskine@mweb.co.za

Abstract

Background: Medical journals use Twitter to engage and disseminate their research articles and implement a range of strategies to maximize reach and impact.

Objective: This study aims to systematically review the literature to synthesize and describe the different Twitter strategies used by medical journals and their effectiveness on journal impact and readership metrics.

Methods: A systematic search of the literature before February 2020 in four electronic databases (PubMed, Web of Science, Scopus, and ScienceDirect) was conducted. Articles were reviewed using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines.

Results: The search identified 44 original research studies that evaluated Twitter strategies implemented by medical journals and analyzed the relationship between Twitter metrics and alternative and citation-based metrics. The key findings suggest that promoting publications on Twitter improves citation-based and alternative metrics for academic medical journals. Moreover, implementing different Twitter strategies maximizes the amount of attention that publications and journals receive. The four key Twitter strategies implemented by many medical journals are tweeting the title and link of the article, infographics, podcasts, and hosting monthly internet-based journal clubs. Each strategy was successful in promoting the publications. However, different metrics were used to measure success.

Conclusions: Four key Twitter strategies are implemented by medical journals: tweeting the title and link of the article, infographics, podcasts, and hosting monthly internet-based journal clubs. In this review, each strategy was successful in promoting publications but used different metrics to measure success. Thus, it is difficult to conclude which strategy is most effective. In addition, the four strategies have different costs and effects on dissemination and readership. We recommend that journals and researchers incorporate a combination of Twitter strategies to maximize research impact and capture audiences with a variety of learning methods.

(J Med Internet Res 2021;23(7):e26378) doi:10.2196/26378

KEYWORDS

Twitter; social media; medical journals; impact

XSL•FO
RenderX

Introduction

The main goals of health science research are to improve health, services, and practice, as well as develop health care technologies [1]. Research needs to be translated from *what we know* to *what we do* to achieve these goals [1]. This process is generally referred to as knowledge translation [2]. Although knowledge translation does not have an agreed-upon definition [3], it typically includes all the steps from the creation of new knowledge to its application [4]. One primary step in knowledge translation is dissemination, that is, the communication and sharing of research findings [1]. Medical journals are a key source of information for health researchers and practitioners. They also play an important role in dissemination [5]. Traditional journal dissemination is passive, unplanned, and uncontrolled [1] and relies on the end user to search for information (known as the *pull* concept). This approach requires awareness of and access to the journal, which is not always the case for health researchers and practitioners [6].

Twitter is a microblogging service that allows the sharing of short messages (tweets) within a 280-character limit and can include images, videos, and hyperlinks to other sites. Users can share the tweet (retweet) with their web-based network or community (followers) [7].

Twitter's global popularity and ubiquitous nature offers a rapid, accessible, and cost-effective medium for communication and sharing of information [8]. Among health professionals and researchers, Twitter is increasingly being used as a medium for health communication to stakeholders [9]. Medical journals have also recognized the potential of Twitter to disseminate their research articles and have implemented a range of strategies to maximize reach [10]. The simplest strategy is posting the title of the research article with a link to the article on the journal's website. Beyond this, medical journals have also summarized research findings into a single post; produced infographics, visual abstracts, and podcasts; and hosted Twitter discussions and journal clubs [10]. In contrast to traditional dissemination, this approach can be considered active, as the journal *pushes* content to its audience [1].

Increased dissemination of research on Twitter can improve article readership and the impact of a journal in terms of citation-based and alternative metrics [10], such as impact factors [11] and citation counts [12]. Research dissemination on Twitter is measured using alternative metrics (*Altmetrics*) and includes Altmetric attention scores, pageviews, article downloads, and Twitter metrics. Altmetrics measure how often an article has been shared, viewed, or referenced on the web by both professional and lay audiences [13]. Compared with citation-based metrics, Altmetrics provide detailed and real-time feedback on the web-based reach and impact of a research article [13]. Despite all the professional benefits it has to offer, Twitter is still an underused tool among medical journals, with less than a third hosting a Twitter profile [14]. One of the main reasons for this underutilization is the lack of evidence-based best practices [10,15]. Thus, studies have been conducted on the effects of different Twitter strategies on article dissemination and journal impact. However, this research has not been synthesized in a meaningful way to inform practice. Therefore, the purpose of this review is to synthesize and describe the different Twitter strategies used by medical journals and their effectiveness on article dissemination, readership, and journal impact metrics.

Methods

Overview

For this review, Twitter was defined as "a microblogging and social networking service on which users post and interact with messages." Medical journals were defined as "a peer-reviewed scientific journal that communicates medical information to health practitioners." Journal impact was measured using citation-based metrics, such as citation count and impact factor. Article readership was measured using pageviews and full-text article downloads. Article dissemination was measured using Altmetric attention scores and Twitter metrics, including impressions, engagements, link clicks, and retweets. The definitions of each dissemination metric are provided in Textbox 1.

XSL•FO
RenderX

Textbox 1. Definitions of outcome measures reported by included studies.

Retweets

- The number of times a user retweeted (reposted) a tweet. This feature allows users to share information from the source with their followers. You can retweet your Tweets or Tweets from someone else [16].

Impressions

- The number of people who saw your tweet [17].

Engagements

- The total number of times a user interacted with a Tweet. Clicks anywhere on the Tweet, including retweets, replies, follows likes, links, cards, hashtags, embedded media, username, profile photo, or Tweet expansion [17].

Link Clicks

- The number of times the article link in the Tweet is clicked [17].

Downloads

- The number of times a research article was download as a pdf from the website of the medical journal [18].

Pageviews

- The number of times an article on the website of the medical journal was loaded on a web browser (abstract and full text combined) irrespective of the source [18].

Altmetric Score

- A weighted count of all the web-based attention an individual research output has received from web-based media platforms, including social media networks, news outlets, blogs, and others [19].

Literature Search Strategy

A systematic review was conducted to retrieve all relevant research studies. All study designs were included in the review to identify the best evidence available to address the research objectives. The literature search was conducted using the following four electronic databases: PubMed, Web of Science, Scopus, and ScienceDirect. The search was performed using the following search terms: “MEDICAL,” “MEDICINE,” “JOURNALS,” “SOCIAL MEDIA,” and “TWITTER.” For example, we searched PubMed using the following strategy: (“medical” OR “medicine”) AND “journals” AND “social media” AND “Twitter.”

Article Selection

Following the collection of studies from different electronic databases, duplicate studies were removed and screened for eligibility using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The inclusion criteria were as follows: (1) English language, full-text original research articles; (2) studies of medical journals; (3) evaluations of Twitter strategies implemented for article dissemination (eg, infographics and web-based journal clubs) or analysis of the relationship between alternative metrics and citation-based metrics; (4) published any time before February 2020; and (5) all research study designs. The exclusion criteria were (1) studies not in English; (2) literature reviews, dissertation theses, review papers, reports, conference papers or abstracts, letters to the editor, commentaries, and feature articles; and (3) studies involving nonmedical academic journals. Following abstract and title screening, we conducted a full-text review of the selected articles.

Data Extraction, Synthesis, and Evaluation

The following data were extracted from the included studies: first author, year of publication, study aim, study sample, Twitter strategy implemented, study methodology (description of the Twitter strategy implemented or metrics being analyzed), primary study outcomes, and results. A customized data extraction sheet was developed using Microsoft Excel. The Medical Education Research Study Quality Instrument was used to evaluate the quality of all included studies [20].

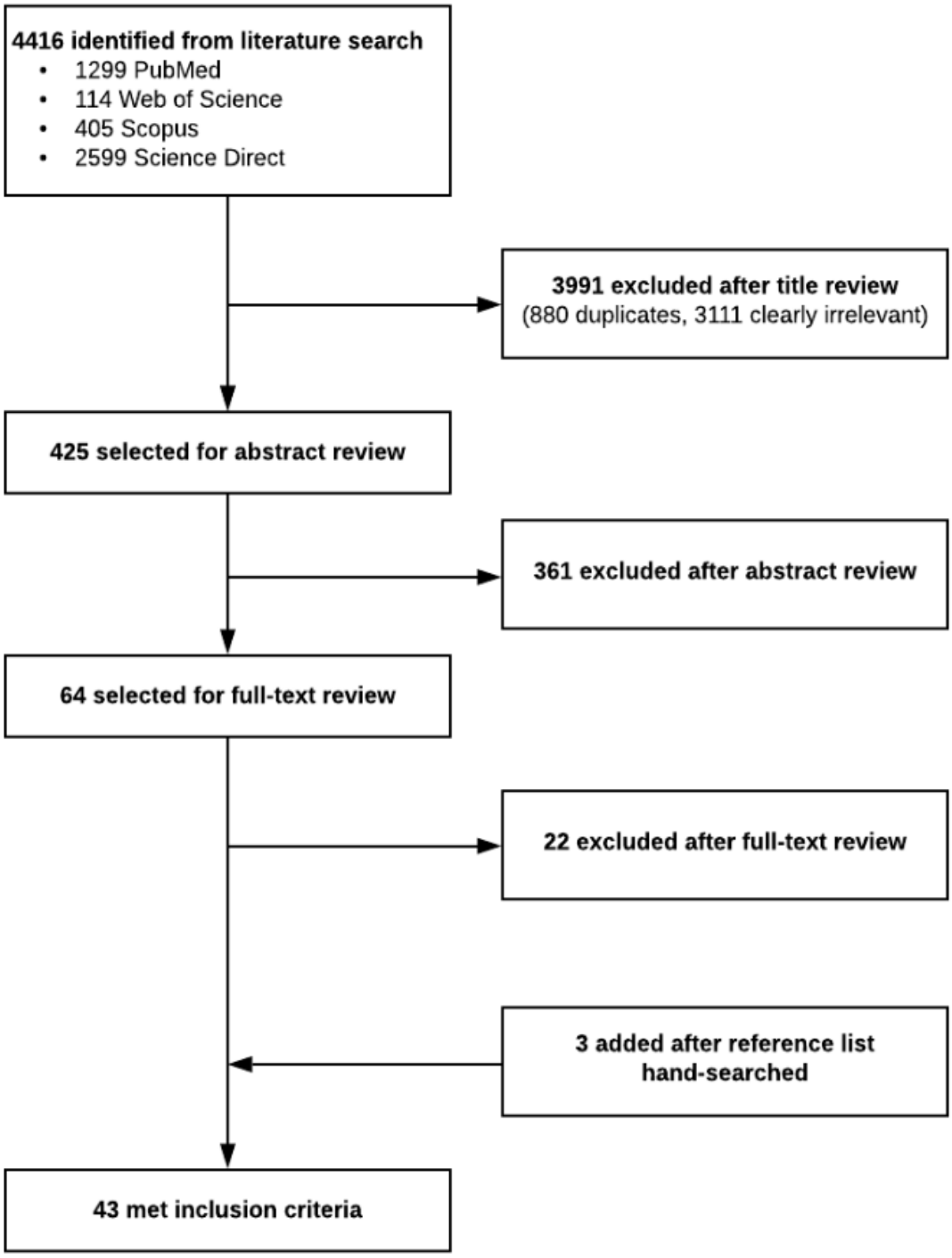
Results

Overview

The database search identified 4416 titles (Figure 1). After removing the duplicates and reviewing the titles, 425 articles remained for abstract review. After reviewing the abstracts, we retrieved 64 full-text articles. The reference list of the 64 full-text articles were manually searched. The total number of included studies was 44. Multimedia Appendix 1 provides a summary of the included studies.

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses flow chart for search strategy.



Characteristics of Included Studies

Of the 43 included studies, 27 (63%) health-related disciplines were reviewed, with urology (6/43, 14%) and radiology (5/43, 9%) representing the top two disciplines. A total of 74% (32/43) of the published studies were within the last 4 years. A total of 53% (23/43) of the studies analyzed the relationship between alternative and citation-based metrics as their primary objective. The remaining 47% (20/43) studied how specific Twitter strategies used by medical journals impacted these relationships. In all, 74% (32/43) of studies used an observational study design [12,14,21-43,45-50] and 26% (11/43) used an experimental design [11,18,44,51-58]. Multimedia Appendix 1 presents the key findings of each study.

Assessing the Quality of Studies

The studies received an average score of 9.5 out of 13.5, ranging from 8 to 10.5. A total of 70% (30/43) of the studies scored above 9.5. For detailed scores, see Multimedia Appendix 2 [10-14,18,21-54,56-59].

XSL•FO
RenderX

Journals With a Twitter Account Have Greater Citation-Based and Alternative Metric Scores

Studies that compared journals with and without a Twitter account showed that Twitter has a higher impact factor (n=4) [27,29,32,34], an increase in Altmetric attention scores (n=1) [39], H-index scores (n=3) [21,31,35], SCImago Journal Rank (n=4) [21,31,35,40], and receive more citations (n=1) and tweets [12].

Number of Tweets May Be Positively Correlated to Publication Citations

A total of 6 studies [22,25,26,28,37,38] reported a positive and significant relationship between Twitter mentions and article citations, whereas 2 studies reported no significant relationship [57,58].

Implementing Twitter Strategies Increases Research Dissemination

Twitter strategies implemented by medical journals to promote their research articles ranged from hosting internet-based journal clubs (7/43, 16%) [42-44,46-50], standard article promotion (7/43, 16%) [11,44,52,53,55,57,58], infographics or visual abstracts (6/43, 14%) [18,45,50,51,54,56], and podcasts (1/43, 2%) [56]. The effects of each strategy are shown in Multimedia Appendix 1, and the definitions for each strategy are provided in Textbox 2. The primary outcome measures included impressions, engagements, link clicks, pageviews (full-text or abstract or both), full-text downloads, and Altmetric scores. The definitions of each outcome are presented in Textbox 1.

Textbox 2. Definitions of the Twitter strategies implemented by journals.

Journal Club

- Participants discuss a selected research article virtually using Twitter, often meeting at a set date and time.

Basic Tweet

- Posting a message on Twitter with the title of an academic article along with a link to the full-text version on the journals' websites. The article may or may not be accessible.

Infographic

- Posting a message on Twitter with the title of an academic article, a summary of the results in the form of an infographic, and a link to the full-text version on the journals' websites. The article may or may not be accessible.

Podcast

- Posting a message on Twitter containing the title of an academic article and a link to the full-text version of the article as well as a downloadable podcast based on the article on the journals' websites. The article may or may not be accessible.

Observational

- No intervention was implemented. The authors analyzed the relationship between Twitter metrics or alternative metrics and citation-based metrics for journals or articles.

Discussion

Principal Findings

This is the first review to analyze the use of Twitter in medical journals and its effect on research dissemination, readership, and journal impact metrics. The first key finding of this review is that journals should endeavor to have a dedicated Twitter account to promote their publications. This Twitter promotion will improve article dissemination, article readership, and citation-based metrics, ultimately leading to an increase in the impact factor. Beyond this, implementing different Twitter strategies maximizes the web-based dissemination of and engagement with publications and journals. There were four main Twitter strategies implemented by medical journals: basic tweeting, infographics, podcasts, and internet-based Twitter journal clubs. Each strategy was successful in the dissemination of publications, but different metrics were used to measure success. Thus, it was difficult to conclude which strategy was the most effective. The discussion below details the benefits and challenges of each strategy and how they increase the potential for successful dissemination.

Basic Tweeting

Basic tweeting was the simplest Twitter strategy used by medical journals, that is, tweeting the title of the article along with the link to the abstract or full-text version on the journal's website. Basic tweeting had varying impacts on article readership and dissemination, based on a number of factors. These factors include the source of the tweet, posting frequency, and the number of Twitter followers.

The tweet source is as important as the content of the tweet. For example, Hawkins et al [44] increased weekly pageviews per article by 139% when the editorial board tweeted articles from their personal accounts compared with articles tweeted from the journal account only. In another study, using a pre-post study design on Twitter, a journal website received a 273% increase in monthly pageviews after a team of physicians and medical graduates (who were active social media users) began tweeting articles from their personal accounts [11]. Similarly, Luc et al [55] showed that tweeting journal articles via the personal accounts of key opinion leaders significantly increased 7-day posttweet Altmetric scores, Mendeley reads, and Twitter impressions. On the basis of these results, we recommend that

XSL•FO
RenderX

journals should encourage key opinion leaders, editorial board members, and active social media users in their medical disciplines to promote articles on their personal Twitter accounts on behalf of the journal.

Increasing the posting frequency may increase the dissemination of articles. Fox et al [52] found no increase in pageviews for articles posted only twice. In contrast, Widmer et al [58] found a 900% increase in pageviews via Twitter for articles posted seven times. In another study, Hawkins et al [44] found that 4 tweets per article resulted in significantly more pageviews than 1 tweet per article. Although the studies differed in design, collectively, these results suggest that posting frequency plays an important role in the effectiveness of the basic tweet strategy and that a dose-response relationship between tweeting and impact may exist [58].

A large number of Twitter followers are essential for the success of social media promotion. Two studies, both by Fox et al [52,53], found no increase in monthly article pageviews after implementing a low-intensity (two posts per article) and high-intensity (three posts per article) posting frequency on the journal's Twitter account. The journal accounts only had 2219 and 10,072 Twitter followers for each study. In contrast, Luc et al [55] and Widmer et al [58] had 52,983 and 1,177,514 Twitter followers, respectively, at the start of their intervention. From a social media recommendation for the journals' perspective, this means journals starting out on Twitter should focus on increasing followers as opposed to being overly concerned about readership metrics.

Basic article promotion is an easy, cost-effective, and time-effective strategy for journals. From a reader's perspective, it keeps them up to date with the latest publications, especially if they choose to be notified every time a journal tweets. However, it requires readers to click the link to the journal's website to read the abstract of the article [51]. In addition, nonscientific readers may not be able to read paid-access articles and may perceive a text abstract as confusing and boring. Clicking the link, however, is beneficial for the journal as it increases their web traffic.

Visual Abstracts

Including a visual abstract, that is, a simplified graphical summary of a study's scientific abstract, within article tweets increases their dissemination on Twitter and, subsequently, readership. In contrast to basic tweets, the visual abstract tweets receive significantly more abstract pageviews, impressions [45,51,54], engagements [45,51], and Altmetric attention scores [56]. However, visual abstracts enhance the impact of basic tweets, which is referred to as the *spillover effect*. For example, two studies reported an increase in impressions, pageviews and engagements for basic tweets sent from the same journal account after implementing the visual abstract strategy [45,54].

Thoma et al [56] and Huang et al [18] found that visual abstracts did not increase full-text readership. Huang et al [18] provided two suggestions for this finding. First, readers may have felt that the visual abstract was a sufficient summary and did not need to read the full-text article [18]. This suggestion justifies the lower number of link clicks received for visual abstract tweets than for basic tweets. Second, readers may not have had access to full-text articles [18]. If journals choose to provide visual abstracts, they need to ensure that the information presented is accurate and easily digestible to prevent misinterpretation and the spread of misinformation.

The responsibility of designing visual abstracts is usually left to the journal editors, which is costly and time-consuming, even though open-source templates are available to help authors create their own [45]. The positive impact of visual abstracts provides a motive for the feasibility of this strategy and for journals to acquire the required resources. Visual abstracts provide an engaging and digestible summary of the research, which is beneficial to nonscientific readers who may not have access to full-text articles. As such, we recommend that journals allocate resources to the creation of visual abstracts.

Podcast

Articles tweeted with linked podcasts receive greater increases in Altmetric attention scores and abstract pageviews than visual abstracts or basic tweets, although they show no increase in full-text readership [56]. A possible explanation for this is that during the podcast, the research paper is critically analyzed and discussed, eliminating the need for users to read the full publication themselves [60]. Podcasts are convenient and accessible and can be heard while performing other tasks, such as commuting, household tasks, or exercise [61]. Podcasts are also enjoyable to listen to due to their entertainment and educational value, and practitioners have reported a positive impact of podcasts on their practices [60]. Creating a podcast is relatively inexpensive, with the minimum required piece of equipment being a dedicated microphone [61].

Internet-Based Twitter Journal Clubs

Hosting monthly 24-hour Twitter discussions of recent publications demonstrate positive growth metrics over the course of the intervention in terms of active participation (number of Twitter profiles involved in the discussion), tweet volume, engagement, and impressions and has the potential to increase traffic to the journal's website [59]. Unlike other Twitter strategies, internet-based journal clubs are based on conversations and real-time discussions [42]. In addition to discussing the study, internet-based journal clubs build relationships between the journal and the audience [49]. Chai et al [42] questioned whether participation in internet-based journal clubs improves knowledge of the topic and whether lessons learned are applied in real-world settings. Another caveat of internet-based journal clubs is that data are only captured on hashtags. As a consequence, important discussion points in tweets without the dedicated journal club hashtags may have been missed [48]. Successful implementation of a Twitter journal club also requires significant preparation and effort. For example, the development of a working group to select topics or manuscripts for tweet chats and having designated moderators to coordinate tweet chats and promote them [41]. If journals choose to implement the internet-based journal clubs strategy, recommendations to increase the likelihood of successful implementation thereof are ensuring that the authors of the studies participate in the web-based discussion [52], allowing opportunities for participants to provide feedback after the

XSL•FO
RenderX

internet-based journal clubs [44], and choosing an optimal time to schedule the chats to allow participants from different time zones to join [52].

Limitations and Future Research

The research designs of studies that tested the effectiveness of the same strategy differed. This made it difficult to draw definitive conclusions regarding Twitter strategy effectiveness [58]. Intervention periods for experimental studies were generally short (7-60 days), and it is unclear whether longer interventions will have greater impacts. In addition, although studies aimed at measuring the impact of different Twitter strategies on outcome measures, some studies also active or promoted articles on other social media platforms at the same time, thereby confounding the results. In terms of the systematic review itself, we tried to focus the search to Twitter studies only. As such, our Boolean terms were "social media" AND "Twitter" as opposed to "social media" OR "Twitter." Admittedly, this may have excluded potential studies and may be considered a caveat of this review. Another potential limitation of our search strategy was the exclusion of the word *tweet* in the search terms.

Conclusions

Twitter is a valuable science communication and marketing tool for academic journals to increase web-based visibility, promote research, and translate science to lay and scientific audiences. Four key Twitter strategies are implemented by medical journals: tweeting the title and link of the article, infographics, podcasts, and hosting monthly internet-based journal clubs. In this review, each strategy was successful in promoting publications but used different metrics to measure success. Thus, it is difficult to conclude which strategy is most effective. In addition, the four strategies have different costs and effects on dissemination and readership. We recommend that journals and researchers incorporate a combination of Twitter strategies to maximize research impact and capture audiences with a variety of learning methods.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of included studies.
[DOCX File , 52 KB - jmir_v23i7e26378_app1.docx]

Multimedia Appendix 2
Quality of included studies using the Medical Education Quality Instruments.
[DOCX File , 25 KB - jmir_v23i7e26378_app2.docx]

References

1. Strauss SE, Tetroe J, Graham ID. Knowledge Translation in Health Care: Moving from Evidence to Practice. Hoboken, New Jersey, United States: John Wiley & Sons; 2013:1-391.
2. Graham I, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
3. McKibbon KA, Lokker C, Wilczynski NL, Ciliska D, Dobbins M, Davis DA, et al. A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: a Tower of Babel? Implement Sci 2010 Feb 12;5(1):16 [FREE Full text] [doi: 10.1186/1748-5908-5-16] [Medline: 21080976]
4. Sudsawad P. Knowledge Translation: Introduction to Models, Strategies and, Measures. National Center for the Dissemination of Disability Research (NCDDR), Southwest Educational Development Laboratory. 2007. URL: https://ktdrr.org/ktlibrary/articles_pubs/ktmodels/ [accessed 2021-07-09]
5. Webber P. The function of questions in different medical journal genres. English Specif Purp 1994 Jan;13(3):257-268. [doi: 10.1016/0889-4906(94)90005-1]
6. Allen HG, Stanton TR, Di Pietro F, Moseley GL. Social media release increases dissemination of original articles in the clinical pain sciences. PLoS One 2013 Jul 17;8(7):e68914 [FREE Full text] [doi: 10.1371/journal.pone.0068914] [Medline: 23874810]
7. Power A. Twitter's potential to enhance professional networking. Br J Midwifery 2015 Jan 02;23(1):65-67. [doi: 10.12968/bjom.2015.23.1.65]
8. Hendricks S, Jones A. European Journal of Sport Science gears up its social media. Eur J Sport Sci 2014;14(6):519-520. [doi: 10.1080/17461391.2014.949586] [Medline: 25141183]
9. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
10. Trueger NS. Medical journals in the age of ubiquitous social media. J Am Coll Radiol 2018 Jan;15(1 Pt B):173-176. [doi: 10.1016/j.jacr.2017.09.036] [Medline: 29122506]

XSL•FO
RenderX

11. Trueger NS, Bokarius AV, Carroll S, April MD, Thoma B. Impact of a physician-led social media sharing program on a medical journal's web traffic. J Am Coll Radiol 2018 Jan;15(1 Pt B):184-189. [doi: 10.1016/j.jacr.2017.09.035] [Medline: 29122507]
12. Ortega JL. The presence of academic journals on Twitter and its relationship with dissemination (tweets) and research impact (citations). Aslib J Info Manag 2017 Nov 20;69(6):674-687. [doi: 10.1108/ajim-02-2017-0055]
13. Trost MJ, Webber EC, Wilson KM. Getting the Word Out: Disseminating scholarly work in the technology age. Acad Pediatr 2017 Apr;17(3):223-224. [doi: 10.1016/j.acap.2017.01.007] [Medline: 28126611]
14. Cosco TD. Medical journals, impact and social media: an ecological study of the Twittersphere. Can Med Assoc J 2015 Dec 08;187(18):1353-1357 [FREE Full text] [doi: 10.1503/cmaj.150976] [Medline: 26644544]
15. Lopez M, Chan TM, Thoma B, Arora VM, Trueger NS. The social media editor at medical journals. Acad Med 2019;94(5):701-707. [doi: 10.1097/acm.0000000000002496]
16. About different types of tweets. Twitter Help Center. URL: https://help.twitter.com/en/using-twitter/types-of-tweets [accessed 2021-07-09]
17. About your activity dashboard. Twitter Help Center. URL: https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard [accessed 2021-07-09]
18. Huang S, Martin LJ, Yeh CH, Chin A, Murray H, Sanderson WB, et al. The effect of an infographic promotion on research dissemination and readership: a randomized controlled trial. Can J Emerg Med 2018 Nov 05;20(6):826-833. [doi: 10.1017/cem.2018.436] [Medline: 30289098]
19. Altmetric. URL: https://help.altmetric.com/support/solutions/articles/6000059309-about-altmetric-and-the-altmetric-score [accessed 2021-07-09]
20. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. J Gen Intern Med 2008 Jul 10;23(7):903-907 [FREE Full text] [doi: 10.1007/s11606-008-0664-3] [Medline: 18612715]
21. Alotaibi N, Guha D, Fallah A, Aldakkan AC, Nassiri F, Badhiwala JH, et al. Social media metrics and bibliometric profiles of neurosurgical departments and journals: is there a relationship? World Neurosurg 2016 Jun;90:574-579. [doi: 10.1016/j.wneu.2016.01.087] [Medline: 26855310]
22. Amath A, Ambacher K, Leddy JJ, Wood TJ, Ramnanan CJ. Comparing alternative and traditional dissemination metrics in medical education. Med Educ 2017 Sep 18;51(9):935-941. [doi: 10.1111/medu.13359] [Medline: 28719136]
23. Chang J, Desai N, Gosain A. Correlation between altmetric score and citations in pediatric surgery core journals. J Surg Res 2019 Nov;243:52-58. [doi: 10.1016/j.jss.2019.05.010] [Medline: 31154133]
24. Dardas LA, Woodward A, Scott J, Xu H, Sawair FA. Measuring the social impact of nursing research: an insight into altmetrics. J Adv Nurs 2019 Jul 24;75(7):1394-1405. [doi: 10.1111/jan.13921] [Medline: 30507052]
25. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
26. Hayon S, Tripathi H, Stormont IM, Dunne MM, Naslund MJ, Siddiqui MM. Twitter mentions and academic citations in the urologic literature. Urology 2019 Jan;123:28-33. [doi: 10.1016/j.urology.2018.08.041] [Medline: 30278190]
27. Hughes H, Hughes A, Murphy C. The use of Twitter by trauma and orthopaedic surgery journals: Twitter activity, impact factor, and alternative metrics. Cureus 2017;9(12):e1931. [doi: 10.7759/cureus.1931]
28. Jeong JW, Kim MJ, Oh H, Jeong S, Kim MH, Cho JR, et al. The impact of social media on citation rates in coloproctology. Colorectal Dis 2019 Oct 19;21(10):1175-1182. [doi: 10.1111/codi.14719] [Medline: 31124259]
29. Kelly B, Redmond CE, Nason GJ, Healy GM, Horgan NA, Heffernan EJ. The use of Twitter by radiology journals: an analysis of Twitter activity and impact factor. J Am Coll Radiol 2016 Nov;13(11):1391-1396. [doi: 10.1016/j.jacr.2016.06.041] [Medline: 27577594]
30. Mullins CH, Boyd CJ, Corey BL. Examining the correlation between altmetric score and citations in the general surgery literature. J Surg Res 2020 Apr;248:159-164. [doi: 10.1016/j.jss.2019.11.008] [Medline: 31901796]
31. Muñoz-Velandia OM, Fernández-Ávila DG, Patino-Hernandez D, Gómez AM. Metrics of activity in social networks are correlated with traditional metrics of scientific impact in endocrinology journals. Diabetes Metab Syndr 2019 Jul;13(4):2437-2440. [doi: 10.1016/j.dsx.2019.06.018] [Medline: 31405656]
32. Nason GJ, O'Kelly F, Kelly ME, Phelan N, Manecksha RP, Lawrentschuk N, et al. The emerging use of Twitter by urological journals. BJU Int 2015 Mar 23;115(3):486-490. [doi: 10.1111/bju.12840] [Medline: 24925047]
33. Nocera AP, Boyd CJ, Boudreau H, Hakim O, Rais-Bahrami S. Examining the correlation between altmetric score and citations in the urology literature. Urology 2019 Dec;134:45-50. [doi: 10.1016/j.urology.2019.09.014] [Medline: 31560915]
34. O'Kelly F, Nason G, Manecksha R, Cascio S, Quinn F, Leonard M, et al. The effect of social media (#SoMe) on journal impact factor and parental awareness in paediatric urology. J Pediatr Urol 2017 Oct;13(5):1-7. [doi: 10.1016/j.jpurol.2017.03.027] [Medline: 28483467]
35. Ordoñez-Gutiérrez JA, Oviedo-Moreno JM, Patino-Hernandez D, Fernández-Ávila DG. Immunology and social networks: an approach towards impact assessment. Rheumatol Int 2020 Feb 12;40(2):251-256. [doi: 10.1007/s00296-019-04459-6] [Medline: 31606774]

XSL•FO
RenderX

36. Patino-Hernandez D, Fernández-Ávila DG, Celis-Preciado CA, Munoz-Velandia OM. Social networks and traditional metrics of impact in pulmonary medicine journals: a correlation study. Adv Respir Med 2020 Jan 14;87(6):209-213. [doi: 10.5603/arm.2019.0058]
37. Rosenkrantz AB, Ayoola A, Singh K, Duszak R. Alternative metrics ("Altmetrics") for assessing article impact in popular general radiology journals. Acad Radiol 2017 Jul;24(7):891-897. [doi: 10.1016/j.acra.2016.11.019] [Medline: 28256440]
38. Smith ZL, Chiang AL, Bowman D, Wallace MB. Longitudinal relationship between social media activity and article citations in the journal Gastrointestinal Endoscopy. Gastrointest Endosc 2019 Jul;90(1):77-83. [doi: 10.1016/j.gie.2019.03.028] [Medline: 30935934]
39. Wang J, Alotaibi NM, Ibrahim GM, Kulkarni AV, Lozano AM. The spectrum of altmetrics in neurosurgery: the top 100 "Trending" articles in neurosurgical journals. World Neurosurg 2017 Jul;103:883-895. [doi: 10.1016/j.wneu.2017.04.157] [Medline: 28478251]
40. Wong K, Piraquive J, Levi JR. Social media presence of otolaryngology journals: the past, present, and future. Laryngoscope 2018 Feb 10;128(2):363-368. [doi: 10.1002/lary.26727] [Medline: 28600839]
41. Zhou JZ, Lemelman BT, Done N, Henderson ML, Macmillan A, Song DH, et al. Social media and the dissemination of research. Plast Reconstr Surg 2018;142(2):555-561. [doi: 10.1097/prs.0000000000004598]
42. Chai PR, Ruha A, Wong KE, Monette DL, Spyres MB, Lapoint J, et al. The virtual toxicology journal club: the dissemination and discussion of noteworthy manuscripts using Twitter. J Med Toxicol 2018 Sep 20;14(3):212-217 [FREE Full text] [doi: 10.1007/s13181-018-0670-8] [Medline: 29926389]
43. Gardhouse AI, Budd L, Yang SY, Wong CL. #GeriMedJC: The Twitter complement to the traditional-format geriatric medicine journal club. J Am Geriatr Soc 2017 Jun 22;65(6):1347-1351. [doi: 10.1111/jgs.14920] [Medline: 28432823]
44. Hawkins CM, Hunter M, Kolenic GE, Carlos RC. Social media and peer-reviewed medical journal readership: a randomized prospective controlled trial. J Am Coll Radiol 2017 May;14(5):596-602. [doi: 10.1016/j.jacr.2016.12.024] [Medline: 28268163]
45. Koo K, Aro T, Pierorazio PM. Impact of social media visual abstracts on research engagement and dissemination in urology. J Urol 2019 Nov;202(5):875-877. [doi: 10.1097/ju.0000000000000391]
46. Luc JG, Ouzounian M, Bender EM, Blitz A, Stamp NL, Varghese TK, et al. The Thoracic Surgery Social Media Network: Early experience and lessons learned. J Thorac Cardiovasc Surg 2019 Oct;158(4):1127-1136. [doi: 10.1016/j.jtcvs.2019.06.099] [Medline: 31422854]
47. McGinnigle E, Francis R, Warriner DR, McAloon CJ. Journal clubs in the digital age: Twitter for continuing professional development. Future Healthc J 2017 Oct 03;4(3):160-166 [FREE Full text] [doi: 10.7861/futurehosp.4-3-160] [Medline: 31098464]
48. Ni hlci T, Archer M, Harrington C, Luc JG, Antonoff MB. Trainee Thoracic Surgery Social Media Network: Early experience with TweetChat-based journal clubs. Ann Thorac Surg 2020 Jan;109(1):285-290. [doi: 10.1016/j.athoracsur.2019.05.083] [Medline: 31348902]
49. Thangasamy IA, Leveridge M, Davies BJ, Finelli A, Stork B, Woo HH. International Urology Journal Club via Twitter: 12-month experience. Eur Urol 2014 Jul;66(1):112-117. [doi: 10.1016/j.eururo.2014.01.034] [Medline: 24548686]
50. Wadhwa V, Latimer E, Chatterjee K, McCarty J, Fitzgerald R. Maximizing the tweet engagement rate in academia: analysis of the AJNR twitter feed. AJNR Am J Neuroradiol 2017 Jun 29;38(10):1866-1868. [doi: 10.3174/ajnr.a5283]
51. Chapman S, Grossman RC, FitzPatrick ME, Brady RR. Randomized controlled trial of plain English and visual abstracts for disseminating surgical research via social media. Br J Surg 2019 Oct 02;106(12):1611-1616. [doi: 10.1002/bjs.11307] [Medline: 31577372]
52. Fox C, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from Circulation. Circulation 2015 Jan 06;131(1):28-33. [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
53. Fox CS, Gurary EB, Ryan J, Bonaca M, Barry K, Loscalzo J, et al. Randomized controlled trial of social media: effect of increased intensity of the intervention. J Am Heart Assoc 2016 May 06;5(5):003088. [doi: 10.1161/jaha.115.003088]
54. Ibrahim A, Lillemoe KD, Klingensmith ME, Dimick JB. Visual abstracts to disseminate research on social media: a prospective, case-control crossover study. Ann Surg 2017 Dec;266(6):46-48. [doi: 10.1097/SLA.0000000000002277] [Medline: 28448382]
55. Luc JG, Archer MA, Arora RC, Bender EM, Blitz A, Cooke DT, et al. Social media improves cardiothoracic surgery literature dissemination: results of a randomized trial. Ann Thorac Surg 2020 Feb;109(2):589-595. [doi: 10.1016/j.athoracsur.2019.06.062] [Medline: 31404547]
56. Thoma B, Murray H, Huang SY, Milne WK, Martin LJ, Bond CM, et al. The impact of social media promotion with infographics and podcasts on research dissemination and readership. Can J Emerg Med 2018 Mar 13;20(2):300-306. [doi: 10.1017/cem.2017.394] [Medline: 28899440]
57. Tonia T, Van Oyen H, Berger A, Schindler C, Künzli N. If I tweet will you cite? The effect of social media exposure of articles on downloads and citations. Int J Public Health 2016 May 18;61(4):513-520. [doi: 10.1007/s00038-016-0831-y] [Medline: 27193574]
58. Widmer RJ, Mandrekar J, Ward A, Aase LA, Lanier WL, Timimi FK, et al. Effect of promotion via social media on access of articles in an academic medical journal. Acad Med 2019;94(10):1546-1553. [doi: 10.1097/acm.0000000000002811]

XSL•FO
RenderX

59. Hawkins CM, Hillman BJ, Carlos RC, Rawson JV, Haines R, Duszak R. The impact of social media on readership of a peer-reviewed medical journal. J Am Coll Radiol 2014 Nov;11(11):1038-1043. [doi: 10.1016/j.jacr.2014.07.029] [Medline: 25439618]
60. Malecki SL, Quinn KL, Zilbert N, Razak F, Ginsburg S, Verma AA, et al. Understanding the use and perceived impact of a medical podcast: qualitative study. JMIR Med Educ 2019 Sep 19;5(2):e12901 [FREE Full text] [doi: 10.2196/12901] [Medline: 31538949]
61. Quintana DS, Heathers JA. How podcasts can benefit scientific communities. Trends Cogn Sci 2021 Jan;25(1):3-5. [doi: 10.1016/j.tics.2020.10.003] [Medline: 33158756]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by R Kukafka, G Eysenbach; submitted 09.12.20; peer-reviewed by P Caldwell, T Chan; comments to author 08.01.21; revised version received 28.02.21; accepted 30.05.21; published 28.07.21.

Please cite as:
Erskine N, Hendricks S
The Use of Twitter by Medical Journals: Systematic Review of the Literature
J Med Internet Res 2021;23(7):e26378
URL: https://www.jmir.org/2021/7/e26378
doi:10.2196/26378
PMID:34319238

©Natalie Erskine, Sharief Hendricks. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Social Media and mHealth Technology for Cancer Screening: Systematic Review and Meta-analysis

Arlinda Ruco[1,2], MPH; Fahima Dossa[3], MD; Jill Tinmouth[1,4,5], MD, PhD, FRCPC; Diego Llovet[1,4], PhD; Jenna Jacobson[1,6], PhD; Teruko Kishibe[7], MISt; Nancy Baxter[1,2,8], MD, PhD, FRCSC

[1]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[2]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada
[3]Division of General Surgery, Department of Surgery, University of Toronto, Toronto, ON, Canada
[4]Prevention & Cancer Control, Ontario Health (Cancer Care Ontario), Toronto, ON, Canada
[5]Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
[6]Ted Rogers School of Management, Ryerson University, Toronto, ON, Canada
[7]Library Services, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada
[8]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Nancy Baxter, MD, PhD, FRCSC
Melbourne School of Population and Global Health
University of Melbourne
207 Bouverie Street
Melbourne, 3053
Australia
Phone: 61 39035551
Email: Nancy.baxter@unimelb.edu.au

Abstract

Background: Cancer is a leading cause of death, and although screening can reduce cancer morbidity and mortality, participation in screening remains suboptimal.

Objective: This systematic review and meta-analysis aims to evaluate the effectiveness of social media and mobile health (mHealth) interventions for cancer screening.

Methods: We searched for randomized controlled trials and quasi-experimental studies of social media and mHealth interventions promoting cancer screening (breast, cervical, colorectal, lung, and prostate cancers) in adults in MEDLINE, Embase, PsycINFO, Scopus, CINAHL, Cochrane Central Register of Controlled Trials, and Communication & Mass Media Complete from January 1, 2000, to July 17, 2020. Two independent reviewers screened the titles, abstracts, and full-text articles and completed the risk of bias assessments. We pooled odds ratios for screening participation using the Mantel-Haenszel method in a random-effects model.

Results: We screened 18,008 records identifying 39 studies (35 mHealth and 4 social media). The types of interventions included peer support (n=1), education or awareness (n=6), reminders (n=13), or mixed (n=19). The overall pooled odds ratio was 1.49 (95% CI 1.31-1.70), with similar effect sizes across cancer types.

Conclusions: Screening programs should consider mHealth interventions because of their promising role in promoting cancer screening participation. Given the limited number of studies identified, further research is needed for social media interventions.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019139615; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=139615

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-035411

(J Med Internet Res 2021;23(7):e26759) doi:10.2196/26759

KEYWORDS

social media; mHealth; cancer screening; digital health; mass screening; mobile phone

XSL•FO
RenderX

Introduction

Background

The use of mobile health (mHealth) technologies and social media in the health care sphere has now become widespread [1-6] and has enabled the rapid sharing of health information, the launching of health promotional campaigns, access to peer support groups, and facilitation of appointment reminders [1,2,4,6]. The World Health Organization has defined mHealth as the use of mobile wireless devices for medical and public health practice [1]. Social media allows those with access to information and communication technology to become content creators and share content with others in virtual communities or networks in addition to accessing information and connecting communities [1,6]. The use of mHealth and social media for health presents an important opportunity to reach health consumers, as these technologies and platforms can provide more frequent interactions, deliver tailored material, and increase accessibility to health information [1], and they now constitute a major way of communicating and advertising. In addition, as access to mobile devices and the internet in low- and middle-resource nations is reported to be comparable with those in developed countries, mHealth and social media may play a role in closing the gap in health disparities between high- and low-resource nations [1,7].

With almost 19 million people expected to be diagnosed with cancer in 2020, cancer is one of the leading causes of death globally [8]. Cancer screening has been shown to reduce disease-specific mortality for a number of cancers [9-12], and as a result, many jurisdictions have implemented population-based screening programs [13,14]. However, screening participation remains suboptimal across jurisdictions and cancer types [13-16]. Emerging research has explored the use of social media and mHealth for cancer screening [17-21]. However, we currently lack an understanding of how effective mHealth and social media can be for cancer screening participation.

Objectives

This systematic review and meta-analysis aims to explore the effectiveness of social media and mHealth interventions to increase cancer screening participation and intention for screen detectable cancers.

Methods

Study Design and Registration

This systematic review was registered with the International PROSPERO (Prospective Register of Systematic Reviews; registration #CRD42019139615) and was written and reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [22].

Inclusion and Exclusion Criteria

Studies included in this systematic review were randomized controlled trials (RCTs) or quasi-experimental studies with a pre- and postintervention design reporting on the effectiveness of an mHealth or social media intervention on cancer screening participation or intention. We included studies pertaining to breast, cervical, colorectal, prostate, or lung cancer, as guidelines for screening exist for these cancers. We defined mHealth interventions as those that delivered health-related information via telecommunication or other wireless technologies (eg, smartphones and tablets) [4]. Social media interventions included those delivered on an already established or new purpose-built social media platform where users could create a profile and share content with other users (virtual communities) [1]. Any comparator was acceptable, including a nonintervention group; an alternate, nonsocial media, non-mHealth intervention; or studies with a pre- and postintervention design. We included studies with multifaceted interventions if at least one component involved a social media– or mHealth-based strategy. Studies were restricted to those conducted in adults aged 18 years or older and articles published in English. In case we were unable to access full-text articles for relevant abstracts, we contacted study authors to obtain the articles. If the authors did not respond, we included the abstract if we could ascertain the eligibility criteria and if the data on the primary or secondary outcome were available. Commentaries, editorials, letters, and reviews were excluded. We also excluded articles published before 2000 because the use of social media was not widespread before this time [4].

Search Strategy

The search strategy was developed by a senior information specialist (TK) and used a combination of text words and MeSH (Medical Subject Headings) terms depending on the database to capture the following concepts: cancer, screening, and social media or mHealth interventions. The search strategy was peer reviewed by a second information specialist in accordance with the Peer Review of Electronic Search Strategies checklist [23] and has been previously published [24].

Information Sources

The search was conducted using the following databases: MEDLINE, Embase, PsycINFO, Scopus, CINAHL, the Cochrane Central Register of Controlled Trials, and Communication & Mass Media Complete from inception to May 31, 2019. The search was updated on July 17, 2020.

Data Management

We used systematic review software (DistillerSR, Evidence Partners Incorporated) to manage records during the screening and study selection phases.

Study Selection

Two independent reviewers (AR and FD) used a piloted data collection form and screened the studies in three stages: title, abstract, and full text. Citations that either reviewer considered potentially eligible at the title stage were included to maximize sensitivity in the early stages of screening. Inclusion in the abstract and full-text screening stages required consensus between the reviewers. Discrepancies between the reviewers at the abstract or full-text stages were resolved by discussion.

Data Extraction

Two reviewers independently extracted data from the included studies using a piloted data collection form in Excel (Version

XSL•FO
RenderX

15.0; Microsoft). Any discrepancies were resolved by discussion. Information extracted from each study included study characteristics (authors, date of publication, location or country, funding, and study design), participant characteristics (sample size, age, sex, ethnicity, and eligibility), intervention details (type of intervention, components, comparator or control group interventions, follow-up or duration, technology platform, and delivery of intervention by whom), and outcomes of interest (screening participation or intention including timeframe).

Outcomes

Screening participation (primary outcome) was defined as the proportion of adults who participated in the screening. This included self-reported outcomes as well as those confirmed through administrative records. Screening intention (secondary outcome) was defined as per the primary study authors. Typically, this is measured as the written intention to undergo screening within a specified timeframe (eg, within the next 3 months or 6 months).

Assessment of Bias

The Cochrane Risk of Bias 2 tool [25] was used to assess the quality of RCTs, and the Cochrane Effective Practice and Organization of Care framework was used to assess bias in pre- and postintervention studies [26]. The risk of bias assessment was independently completed for each study by 2 reviewers (AR and FD). Discrepancies were resolved by discussion or by a third investigator if needed. The *Robvis* tool was used to create a risk of bias plot [27].

Data Synthesis and Analysis

The study, participant, and intervention characteristics and the risk of bias assessments are presented descriptively. We categorized interventions based on their nature, including (1) reminders, (2) education or awareness, (3) navigation or counseling, (4) peer support, (5) decision aids, and (6) mixed. We report on the outcomes of interest in absolute and relative terms and pooled odds ratios (ORs) for screening participation from RCTs using the Mantel-Haenszel method in a random-effects model. If the outcome was measured at several time points, we used the values from the longest follow-up for our study. In RCTs where several intervention arms had a social media or mHealth component, we included them in our analysis and divided the proportion screened of the control or comparison group equally by the number of intervention arms of interest to maintain the same proportion of those screened while not counting the sample size of the control group more than once, as recommended by Cochrane [28]. Forest plots were created to graphically display results stratified by cancer type and the nature of the intervention. Statistical heterogeneity was calculated using the I^2 statistic, where a cutoff of ≥75% was defined as considerable heterogeneity [28]. We conducted a sensitivity analysis in which we excluded articles that were assessed to have a high risk of bias. In addition, we conducted sensitivity analyses to explore whether the overall pooled effect estimate would differ for studies measuring the outcome of cancer screening participation through self-reporting compared with objective or administrative records and for studies conducted in low- and middle-income countries (LMICs). We checked for publication bias for the primary outcome among the RCTs using a funnel plot. Statistical significance was set at a two-tailed $P<.05$. Meta-analyses were performed using Review Manager (RevMan, The Cochrane Collaboration) 5.0.

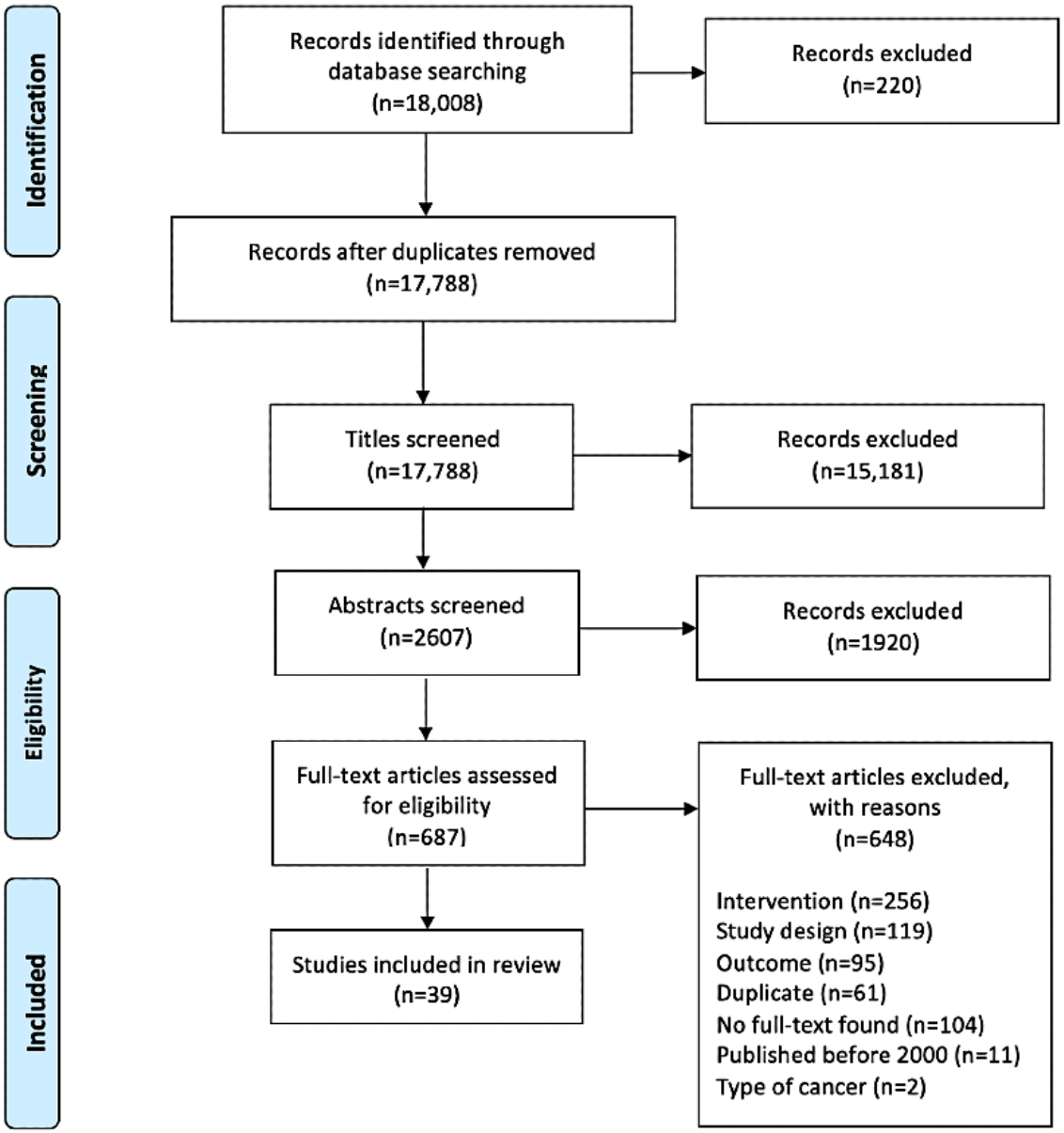
Results

Search Results and Characteristics of Included Studies

A total of 18,008 records were identified in the search. After duplicates were removed, 17,788 titles, 2607 abstracts, and 687 full-text articles were screened. After all the eligibility criteria were applied, 39 articles were included [29-67] (Figure 1). Table 1 presents a summary of the included RCTs (n=30), and Table 2 presents an overview of the included pre- and postintervention studies (n=9). Briefly, the studies that were included were published between 2011 and 2020 and conducted in North America, Europe, Asia, and Africa. Most of the studies (35/39, 90%) described mHealth interventions, and 10% (4/39) of them included social media. The most common type of intervention was mixed (n=19), followed by reminders (n=13), education or awareness (n=6), and peer support (n=1). Mixed interventions were most commonly a combination of reminder and education strategies. There were 16 studies focused on cervical cancer, 14 on colorectal cancer (CRC), 7 on breast cancer, and 1 each on lung and prostate cancer screening. The interventions were implemented by public or private screening programs, university-based research teams, or health care centers or units.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram outlining the steps involved in identifying screened and included studies in the meta-analysis.



XSL•FO
RenderX

Table 1. Summary of included randomized controlled trials (n=30).

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
Arcas et al [29]	Spain	Breast	mHealth[a]	Reminder	703	Women (aged 50-69 years) with a registered mobile phone number	Invitation letter and text message reminder 2 days before the mammography appointment	• Proportion that screened for breast cancer during the 2-month rescreening period
Vidal et al [51]	Barcelona, Spain	Breast	mHealth	Reminder	12,786	Breast cancer screening target population of the southern Barcelona metropolitan area	Text message reminder 3 days before a scheduled appointment with or without a message, with a new appointment date if requested	• Proportion attending an appointment before October 31, 2011 (3-5 months after the intervention)
Kerrison et al [41]	United Kingdom	Breast	mHealth	Reminder	2240	Women (aged 47-53 years) who were due to be invited for their first routine breast screen	Text message reminder 48 hours before the appointment and an additional text message if they did not attend the initial appointment	• Proportion attending the appointment within 60 days of the initial appointment
Rashid et al [47]	Klang, Malaysia	Cervical	mHealth	Reminder	1000	Women (aged 20-65 years) residing in Klang who had a nonpositive Papanicolaou test in the previous year and were due for repeat screening	Text message reminder for a repeat Papanicolaou test within a month from the date of recall	• Proportion completing the Papanicolaou test within 8 weeks
Wanyoro and Kabiru [52]	Thika, Kenya	Cervical	mHealth	Reminder	286	Women (aged 25-70 years) attending the general outpatient clinic who had never had cervical cancer screening, who owned a mobile phone, and who had normal cervical Papanicolaou test after the initial baseline screening	4 text message reminders in a period of 2 weeks	• Proportion screened for cervical cancer at the same site within 2 weeks
Huf et al [39]	United Kingdom	Cervical	mHealth	Reminder	14,587	Women (aged 24-64 years)	1 of 6 text message reminders: a simple reminder, general practice endorsement, total and proportional social norms messages, and gain- and loss-framed messages	• Proportion who screened within 18 weeks after the reminder

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
Sly et al [50]	New York, United States	CRC[b]	mHealth	Reminder	24	Adults (aged >50 years) with referral for screening colonoscopy with no personal or family history of CRC or any chronic gastrointestinal disorder, with telephone service, and who spoke English	Standard navigation, a scheduling telephone call and 2 text message appointment reminders	• Colonoscopy completion within 3 months
Hagoel et al [36]	Israel	CRC	mHealth	Reminder	48,091	Adults (aged 50-74 years) with no diagnosis of an inflammatory bowel disease or a bowel malignancy, who had not undergone colonoscopy within the previous 3 years, and who had not performed FOBT[c] in the previous year	Text message reminders including interrogative or noninterrogative messages	• Proportion completing FOBT at 6 months
Coronado et al [32]	United States	CRC	mHealth	Reminder	2010	Adults (aged 50-75 years) not up to date with CRC screening and with a clinic visit in the previous year	2 text message reminders with or without a live phone call	• FIT[d] kit return rate
Hirst et al [38]	United Kingdom	CRC	mHealth	Reminder	8269	Adults (aged 60-74 years)	Usual care and a text message reminder if they had not returned their test kit within 8 weeks	• Proportion returning test kit at the end of an 18-week screening episode
Lam et al [61]	Hong Kong	CRC	mHealth	Reminder	500	Adults (aged 40-70 years) who were asymptomatic and had a previous negative FIT test and who were expected for an annual FIT screening in the subsequent year	A WhatsApp message reminder sent 1 month before the due date for subsequent FIT	• Proportion successfully returning the FIT kit
Coronado et al [33]	Los Angeles, United States	CRC	mHealth	Reminder	1767	Adults (aged 50-75 years) who were overdue for CRC screening and had attended at least two clinic visits within the past 24 months	Text message prompt before receipt of the FIT kit with 2 automated phone call reminders or with 2 automated phone calls and up to 3 live phone call reminders	• Proportion completing the FIT kit within 6 months
Hwang et al [40]	United States	CRC	Social media	Peer support	306			

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
						Adults (aged 50-75 years) who had no previous diagnosis of CRC, had no history of inflammatory bowel disease, and were not up to date with CRC screening	Study-specific web-based *Spark-Team* to access the narratives and interact with the narrators (positive role models) and other participants	• Proportion screened for CRC at 6 months (FOBT, sigmoidoscopy, or colonoscopy)
Lakkis et al [43]	Beirut, Lebanon	Breast	mHealth	Mixed (education and reminder)	385	Women (aged 40-75 years) who had not undergone a mammogram in the past 2 years	Educational and general invitation text message for mammography and 3 additional text reminders	• Completion of a mammography
Chung et al [31]	Republic of Korea	Breast	mHealth	Mixed (education and reminder)	202	Women (aged 20-65 years) who underwent surgery for breast cancer, excluding those with distant metastasis or recurrent breast cancer	Usual care and 1 text message reminder and 1 educational text message	• Adherent to monthly BSE[e] for 5 out of 6 months
Heydari and Noroozi [37]	Bushehr, Iran	Breast	mHealth	Mixed (education and reminder)	120	Women (aged ≥40 years) who were elementary school teachers, were not pregnant or breastfeeding, had no history of cancer, had no family history of breast cancer, had not had breast biopsy experience and mammography in the past 3 years	Multimedia education session through a CD and text messages; 1-2 educational text messages sent on a weekly basis for 1 month and a reminder about mammography	• Proportion completing mammography • Intention to get a mammography
Lee et al [44]	Minnesota, United States	Breast	mHealth	Mixed (education and navigation)	131	Korean American immigrant women (aged 40-79 years) who had not received a mammogram in the past 2 years	mMammogram mobile app delivering 8-21 messages over a 7-day period	• Proportion receiving mammography or with a scheduled appointment within 6 months • Intention to receive a mammography in the future on a 4-point scale (1=not within a year, 2=within a year, 3=within 3 months, and 4=within 1 month)
Khademolhosseini et al [42]	Bushehr, Iran	Cervical	mHealth	Mixed (education and reminder)	95		Educational training through text messaging, electronic posters, infographics, podcasts, and video tutorial and a reminder to perform a Papanicolaou smear test	• Completion of the Papanicolaou test within 3 months

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
						Women who were able to read and write, were married for at least 6 months, had a smartphone, had no history of genital tract cancer in their family, and had no experience of doing a Papanicolaou smear test in the past 3 years		
Richman et al [49]	North Carolina, United States	Cervical or rectal	mHealth	Mixed (education and reminder)	264	Adults (aged 18-26 years) who attended the university and who were voluntarily initiating the first HPV[f] vaccine dose from the campus student health center	7 electronic email or text messages once per month for 7 months	• Proportion completing HPV dose 3 vaccine
Adler et al [62]	United States	Cervical	mHealth	Mixed (education and reminder)	95	Women (aged 21-65 years) with no past hysterectomy with cervical removal or known HIV infection	Referral and 3 text messages delivered at 30-day intervals over a period of 90 days after enrollment	• Proportion who underwent cervical cancer screening 150 days after enrollment
Erwin et al [34]	Kilimanjaro and Arusha regions, Tanzania	Cervical	mHealth	Mixed (education and reminder)	851	Women (aged 25-49 years) with access to a mobile phone living in the catchment areas of Mawenzi Regional Referral Hospital and Meru District Hospital	15 unique text messages delivered over 21 days with or without a transportation e -voucher covering return transportation to the nearest screening clinic	• Proportion attending cervical cancer screening within 60 days
Firmino-Machado et al [35]	Portugal	Cervical	mHealth	Mixed (education and reminder)	1220	Women (aged 25-49 years) eligible for screening and registered at primary health care units that perform systematic written letter invitations for screening	Automated or customized text messages and phone calls, followed by text message reminders of the appointment (step 1), phone calls by clinical secretaries (step 2), and phone calls or face-to-face interviews by doctors (step 3)	• Proportion adherent to cervical cancer screening at 45 (step 1), 90 (step 1+2), and 150 days after the initial invitation (step 1+2+3)
Linde et al [65]	Tanzania	Cervical	mHealth	Mixed (education and reminder)	689	Women (aged 25-60 years) who had tested positive for HPV during a patient-initiated opportunistic screening 14 months earlier	10 educative text messages (1 per month) and 5 reminders (14, 7, and 1 day before the scheduled screening appointment) over a 10-month period	• Proportion attending the scheduled screening appointment within 30 days
Romli et al [63]	Kedah, Malaysia	Cervical	mHealth		210			

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
				Mixed (education and reminder)		Women entrepreneurs (aged 20-65 years) who received financial help from Amanah Ikhtiar Malaysia and who were or had been previously married	A 30-minute educational talk, a 5-minute video on Papanicolaou smear test procedures, experience sharing from a cervical cancer survivor, distribution of pamphlet on cervical cancer and Papanicolaou smear testing, and 2 text message reminders sent over a 3-month period	• Proportion having a Papanicolaou smear test
Baker et al [30]	Chicago, United States	CRC	mHealth	Mixed (education, reminder, and navigation)	450	Adults (aged 51-75 years) with preferred language listed as English or Spanish and with a negative FOBT	A mailed reminder letter and FIT kit with postage-paid envelope, automated telephone and text message reminders, and personal telephone outreach by a screening navigator after 3 months	• Proportion completing either FOBT or colonoscopy within 6 months of the date the patient was due for annual screening
Muller et al [46]	Anchorage, Alaska	CRC	mHealth	Mixed (education and reminder)	2386	Alaska Native or American Indian adults (aged 40-75 years) with no history of CRC or colectomy enrolled with the Southcentral Foundation health care system and eligible for screening	A maximum of 3 text messages over 2 months	• Proportion screened (FIT, FOBT, flexible sigmoidoscopy, or colonoscopy)
Miller et al [45]	North Carolina, United States	CRC	mHealth	Mixed (education and decision aid)	450	English-speaking adults (aged 50-74 years) who were scheduled to see a primary care provider and were due for CRC screening	mPATH-CRC, an iPad app providing screening information, help with screening decision, *self-ordering* a screening test, and automated electronic messages to complete the chosen test	• CRC screening completed within 24 weeks • Intention to receive screening within the next 6 months
Reiter et al [48]	United States	Rectal	mHealth	Mixed (education and reminder)	150	Gay or bisexual men (aged 18-25 years) residing in the United States who had not received any HPV vaccine doses	Population-targeted, individually tailored content about HPV and monthly HPV vaccination reminders sent via email and/or text message	• Proportion completing all 3 doses of the HPV vaccine
Wong et al [53]	Hong Kong	CRC	mHealth	Mixed (education and reminder)	629			

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
						Adults (aged 40-70 years) at average risk of CRC who had a negative FIT result in their first screening round for the study	Generic text message about the importance of regular CRC screening and the time and venue of FIT tube retrieval	• Proportion successfully returning completed FIT specimen within 6 months
Mahmud et al [64]	United States	CRC	mHealth	Mixed (education and reminder)	71	Adults (aged 18-75 years) scheduled for outpatient colonoscopy within 2 months of initial contact	9 text messages sent in the week before the scheduled procedure	• Proportion who attended their scheduled appointment

[a]mHealth: mobile health.

[b]CRC: colorectal cancer.

[c]FOBT: fecal occult blood test.

[d]FIT: fecal immunochemical test.

[e]BSE. breast self-exam.

[f]HPV: human papilloma virus.

The most common reminder strategies used were text message reminders [29-39,41-43,46-55,57-65]. Educational strategies most commonly included general health information about the specific cancer and information about cancer screening, including the importance of screening. Although text messages were commonly used to deliver educational information [34,35,37,42-44,46,48,49,53-55,59,62,64,65], some studies also used electronic posters or infographics, CDs, videos, mobile apps, and podcasts [37,42,44,45,55,59,63]. Education was also provided through in-person educational or training sessions in some cases in addition to a social media or mHealth strategy or in the comparison groups [55,63]. Educational interventions using social media included social media campaigns [56] or sharing information or daily posts about screening or cancer with participants who were members of a group (virtual community) on a social media platform [66,67]. Peer support interventions on social media also leveraged groups to support participants of that virtual community through the sharing of personal stories and narratives [40]. Outcomes were measured at several time points, including the proportion attending a scheduled appointment or those participating in screening within 2 weeks [52], a month [65], 45 days [35], 60 days [29,34,41,47], 3 months [35,42,50], 3-5 months [38,39,51], or 6 months [30,31,33,36,40,45,53].

There was wide variability in the study participants. For example, the included participants were targeted based on geographical region in some studies [34,51,56] or by their profession as elementary school teachers [37], entrepreneurs [63], or university students [49,59]. Some studies were targeted to specific racial and cultural groups [44,46,54,58,67], whereas others included gay and bisexual men only [48] or women who were HIV positive [60]. The intervention intensity also differed between the studies. For example, some interventions included sending only a single text message reminder [29,31,33,38,39,41,51], whereas others included sending 22 text messages over 16 days [54] or 21 messages over a 7-day period [44]. For social media interventions, participants in one study received three daily posts over a 12-week period [67] or as many as 20 posts per day over 5 days [66].

XSL•FO
RenderX

Table 2. Summary of included pre- and postintervention studies (n=9).

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
Ganta et al [60]	Nevada, United States	Cervical	mHealth[a]	Reminder	473	HIV-infected women (aged ≥18 years) at the HIV Wellness Center	Reminders to schedule a Papanicolaou test via 3 sequential text messages and subsequently by 3 phone call attempts	• Proportion completing the Papanicolaou test
Lee et al [58]	Minnesota, United States	Cervical	mHealth	Education or awareness	30	Korean American women (aged 21-29 years) with no previous receipt of a Papanicolaou test with up-to-date health insurance	7-day text message–based intervention including quizzes and questions and engagement in conversation	• Proportion receiving a Papanicolaou test within 3 months • Intent to receive a Papanicolaou test within a year
Lemos et al [59]	Madeira, Portugal	Cervical	mHealth	Education or awareness	144	Female college students recruited from various undergraduate courses of Madeira University	5 structured text messages delivered over 5 weeks and an educational video intervention lasting 12 minutes	• Intention to get a Papanicolaou test measured on a 5-point Likert scale from 1 (definitely will not do) to 5 (definitely will do)
Le and Holt [54]	United States	Cervical	mHealth	Education or awareness	52	Church-attending African-American women (aged 21-65 years) with no previous medical history of cervical cancer or hysterectomy	22 text messages delivered over 16 days, containing health-specific and spiritually based content	• Intent to get a Papanicolaou smear test in the next 6 months
Lyson et al [66]	United States	Cervical	Social media	Education or awareness	782	Women (aged ≥18 years) who lived in the United States, spoke English as their primary language, and did not have cervical cancer	Health Connect web-based platform where participants were assigned to groups of 9 and where each participant was randomly distributed a set of 20 tweets or messages per day over 5 days in a personalized message feed	• Proportion ever had a Papanicolaou test • Proportion ever received the HPV[b] vaccine
Key et al [67]	Kentucky, United States	CRC[c]	Social media	Education or awareness	60	Appalachian Kentuckians (aged ≥50 years) noncompliant with current screening guidelines	Participants joined a closed Facebook group and were presented with 3 daily Facebook posts during the 12-week intervention	• Proportion ever received a colonoscopy or FOBT[d]

XSL•FO
RenderX

Study	Location	Type of cancer	Intervention type	Nature of intervention	Total sample size	Population	Summary of intervention	Outcomes
Jessup et al [56]	Massachusetts, United States	Lung	Social media	Education or awareness	Variable depending on platform	Patients, caregivers, and health care providers within a 60-mile radius of a large quaternary medical center and 2 affiliated off-campus imaging sites. Patient campaign targeted current and former smokers (aged ≥55 years), females (aged ≥55 years), patients and employees of the academic medical center (aged ≥18 years), and caregivers (aged ≥18 years)	Patient awareness campaign on Facebook and Google and provider campaign on LinkedIn and Twitter	• Number of LDCT[e] examinations per week before and after the campaign
Fornos et al [57]	Texas, United States	Cervical	mHealth	Mixed (education, reminders, and navigation)	32,807	Women (aged ≥18 years) enrolled in CareLink who were not up to date with Papanicolaou screening or actively obtaining Papanicolaou test appointments	Newsletters, public service announcements, automated client reminders including text messages, and community outreach	• 3-year cervical cancer screening rate
Capik and Gozum [55]	Erzurum, Turkey	Prostate	mHealth	Mixed (education and reminders)	75	Men (aged 41-65 years) working in 2 public institutions who had not received a prostate cancer diagnosis	Poster announcements, interactive educational session, access to website, desk calendar information and reminders, monthly email reminders, flyers, and 1 text message	• Proportion having had a PSA[f] test in the last 3 months • Proportion having had a prostate examination in the last 3 months

[a]mHealth: mobile health.

[b]HPV: human papilloma virus.

[c]CRC: colorectal cancer.

[d]FOBT: fecal occult blood test.

[e]LDCT: low-dose computed tomography.

[f]PSA: prostate-specific antigen.

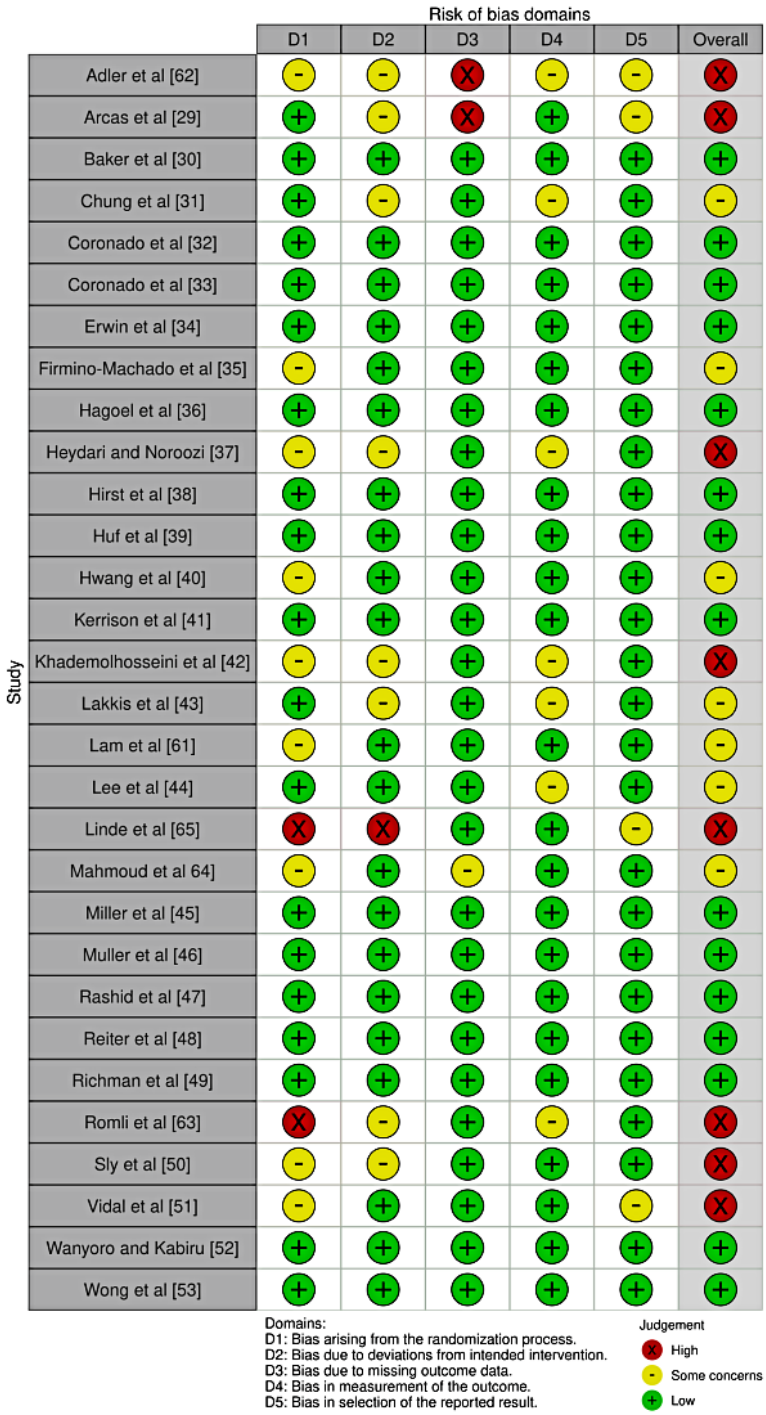
Quality Assessment

Risk of bias assessments for the included studies are shown in Figures 2 and 3. Briefly, 27% (8/30) of the included RCTs were classified as high risk, 23% (7/30) as having some concerns, and the remainder (15/30, 50%) were classified as low risk. Common reasons for being classified as high risk included having some concerns in several domains, including bias arising from the randomization process, effect of assignment to intervention, and measurement of the outcome. All pre- and postintervention studies were classified as high risk. Figure 4 displays the funnel plot used to check for publication bias. The x-axis represents the effect estimates, whereas the y-axis represents the study size or precision. The funnel plot generated

XSL•FO
RenderX

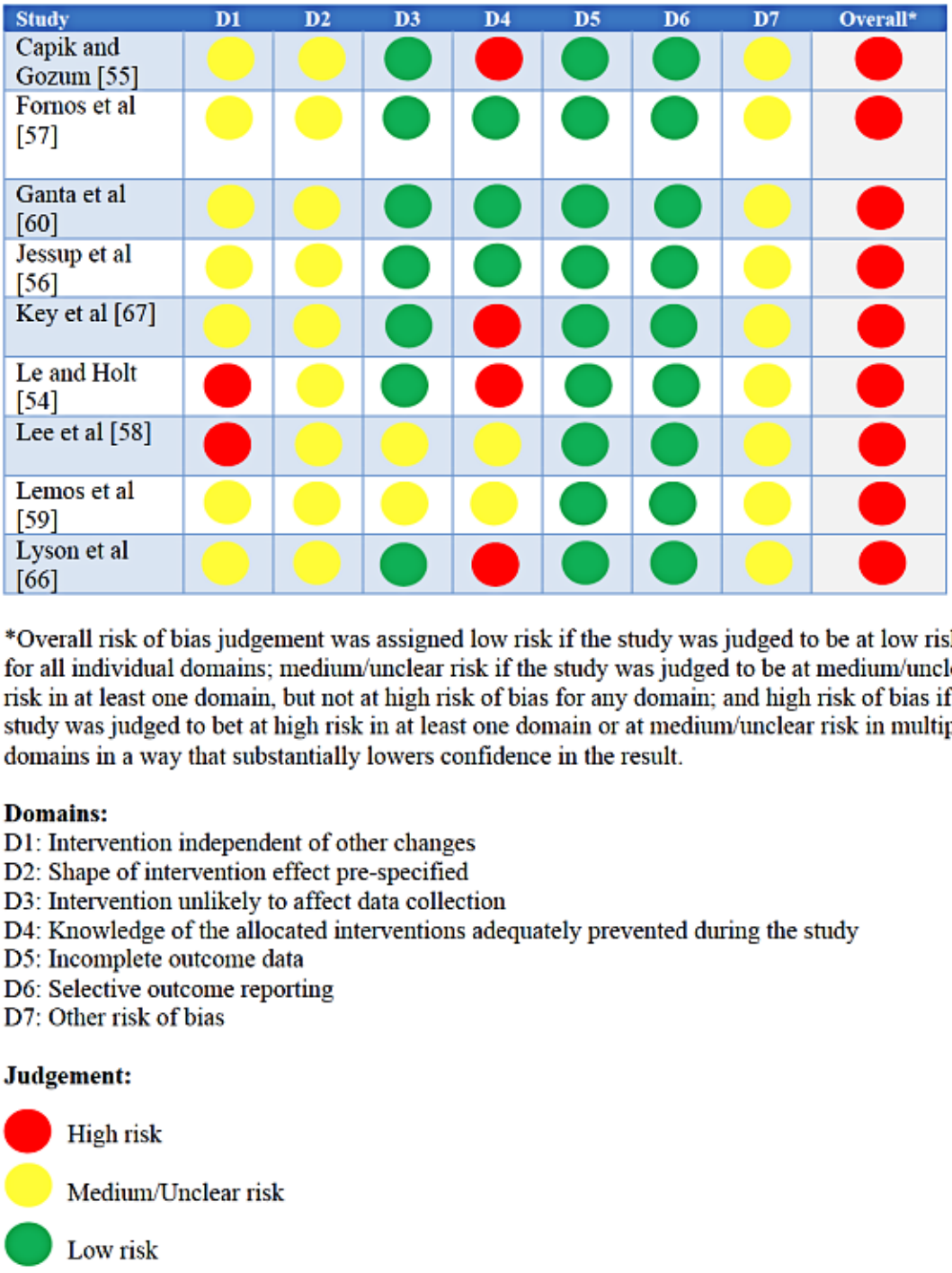
may suggest some publication bias because of the lack of studies in the bottom left corner of the plot representing studies with small effect sizes and variances.

Figure 2. Risk of bias assessment for the included randomized controlled trials (n=30) created using the Robvis tool.



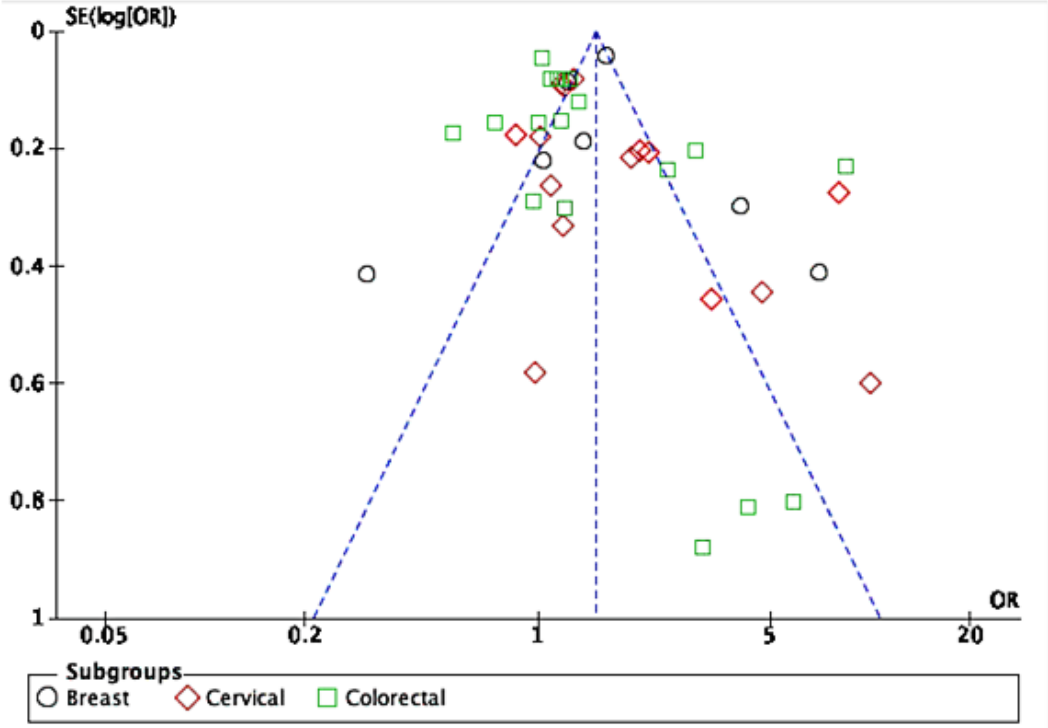
XSL•FO
RenderX

Figure 3. Risk of bias assessment for the included pre- and postintervention studies (n=9).



XSL•FO
RenderX

Figure 4. Funnel plot of publication bias for the randomized controlled trials reporting on the primary outcome. OR: odds ratio.



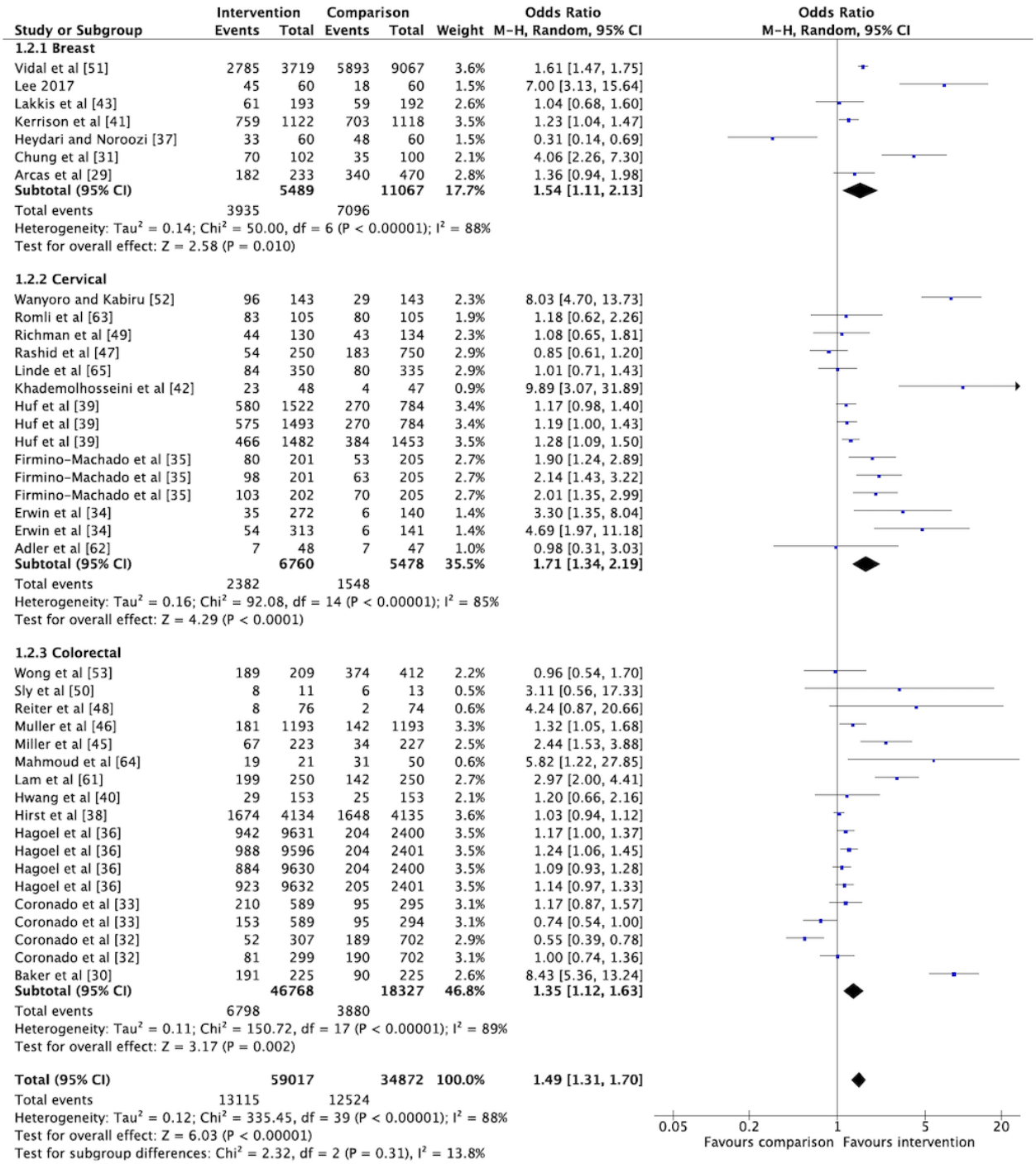
Primary and Secondary Outcomes

The absolute effect of being screened in the intervention arms was 22.22% (13,115/59,017). There was an absolute risk difference of 14% (95% CI 13.12-14.33) between the intervention and comparison arms, with the proportion screened in the comparison arms being 35.94% (12,524/34,872). When stratified by cancer type, the absolute proportion screened in the intervention arms was 71.68% (3935/5489) for breast cancer compared with 64.11% (7096/11,067) in the comparison arms (risk difference 8%; 95% CI 6.08-9.06). For cervical cancer, there were 35.23% (2382/6760) screened in the intervention arms compared with 28.26% (1548/5478) in the comparison arms. For CRC, the proportion screened in the intervention arms was 14.53% (6798/46,768) and 21.17% (3880/18,327) in the comparison arms, with a risk difference of 6% (95% CI 5.96-7.31).

The overall pooled OR for cancer screening participation among the included RCTs was 1.49 (95% CI 1.31-1.70; Figure 5), indicating that the odds of getting screened increased by 49% for those who received a social media or mHealth intervention. However, considerable heterogeneity was observed (I^2=88%). Similar effect estimates were observed when stratified by cancer type, with the largest effect observed for cervical cancer screening studies (OR 1.71, 95% CI 1.34-2.19; Figure 5). Stratification by cancer type did not reduce the heterogeneity. When we conducted a sensitivity analysis excluding trials assessed to have a high risk of bias, the overall pooled OR and I^2 remained stable (OR 1.54, 95% CI 1.33-1.78; Figure 6). The overall pooled OR was not significant when including only studies measuring screening participation through self-reporting (OR 2.09, 95% CI 0.96-4.53). The overall pooled effect estimate remained stable when including only studies that captured the outcome through administrative records (OR 1.46, 95% CI 1.28-1.66). When we included only studies conducted in LMIC settings (n=3), the overall pooled OR was 3.29 (95% CI 1.02-10.60) with considerable heterogeneity (I^2=93%). However, the pooled OR increased to 5.50 (95% CI 3.19-9.51) with only moderate heterogeneity (I^2=38%) when only studies with a low risk of bias were included (n=2). We also conducted subgroup analyses by meta-analyzing studies based on the nature of the intervention. The results showed an overall pooled effect estimate of 1.23 (95% CI 1.08-1.41) for reminder interventions (Figure 7) and 2.07 (95% CI 1.49-5.83) for mixed interventions (Figure 8). Heterogeneity did not change when subgroup analyses were conducted.

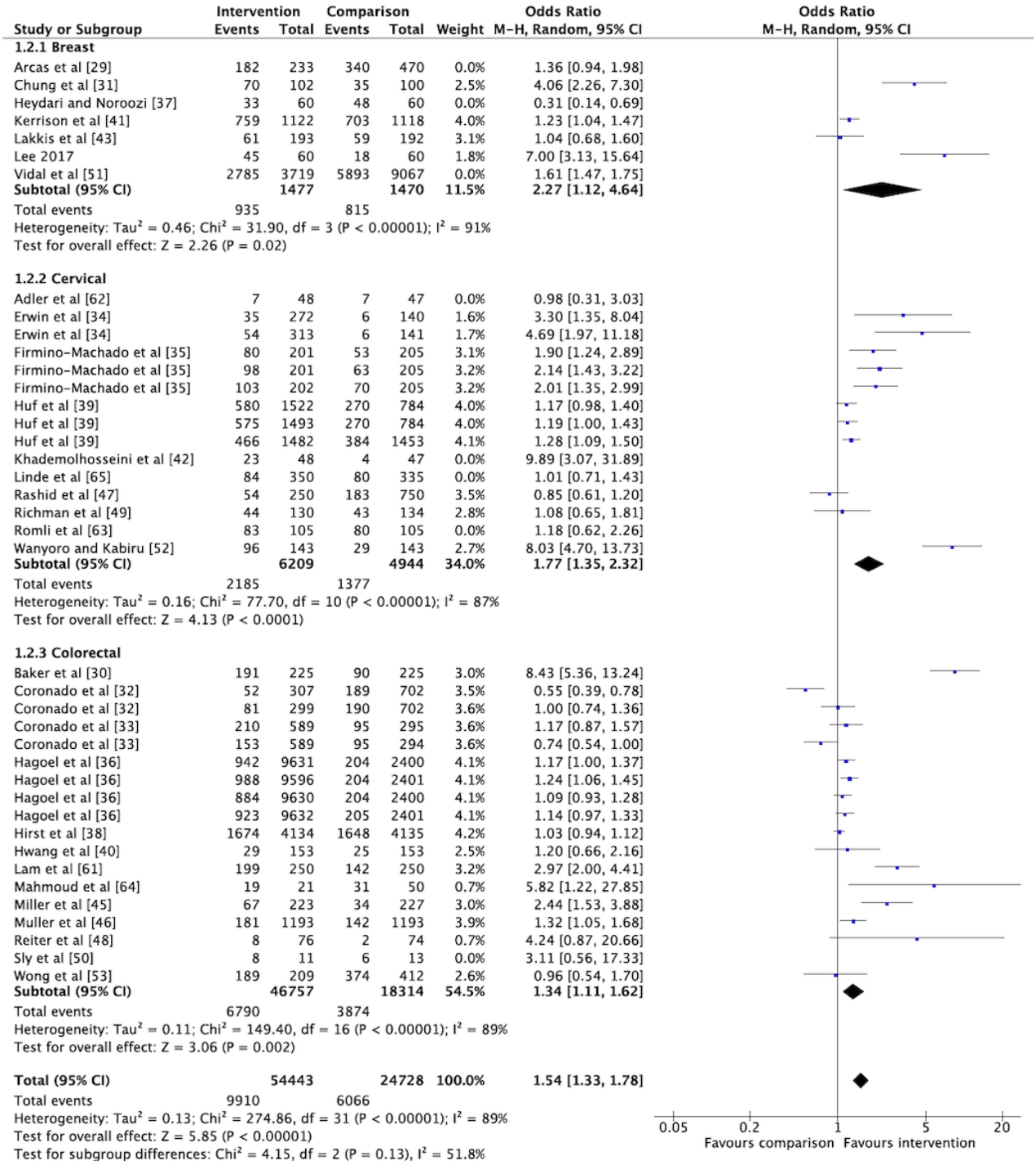
XSL•FO
RenderX

Figure 5. Forest plot for the randomized controlled trials reporting on the primary outcome of cancer screening participation categorized by type of cancer (n=30).



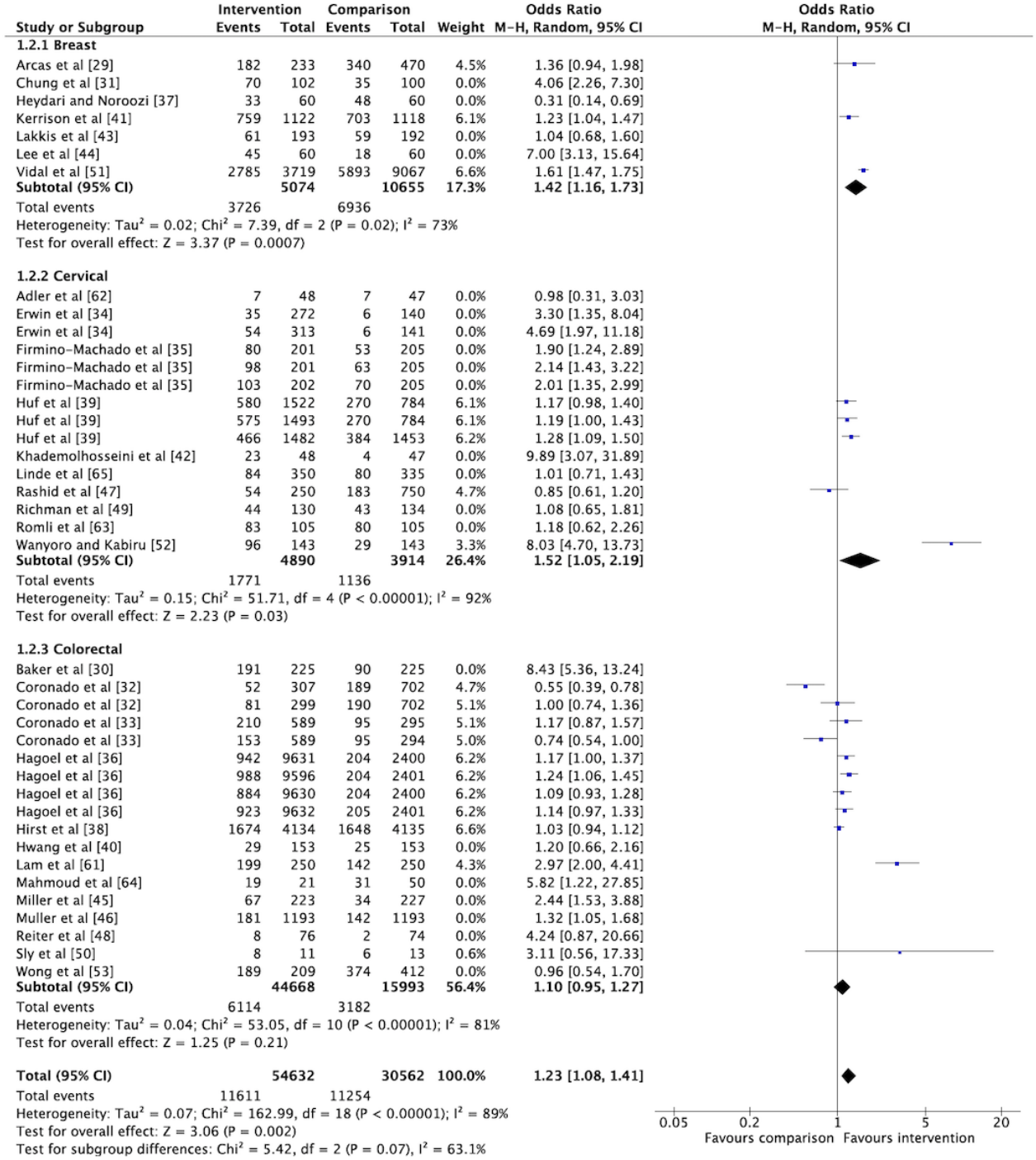
XSL•FO
RenderX

Figure 6. Sensitivity analysis for the primary outcome of interest of cancer screening participation without inclusion of randomized controlled trials with a high risk of bias (n=22).



XSL•FO
RenderX

Figure 7. Forest plot for the reminder interventions reporting on the primary outcome of cancer screening participation (n=12).



XSL•FO
RenderX

Figure 8. Forest plot for the mixed interventions reporting on the primary outcome of cancer screening participation (n=17).

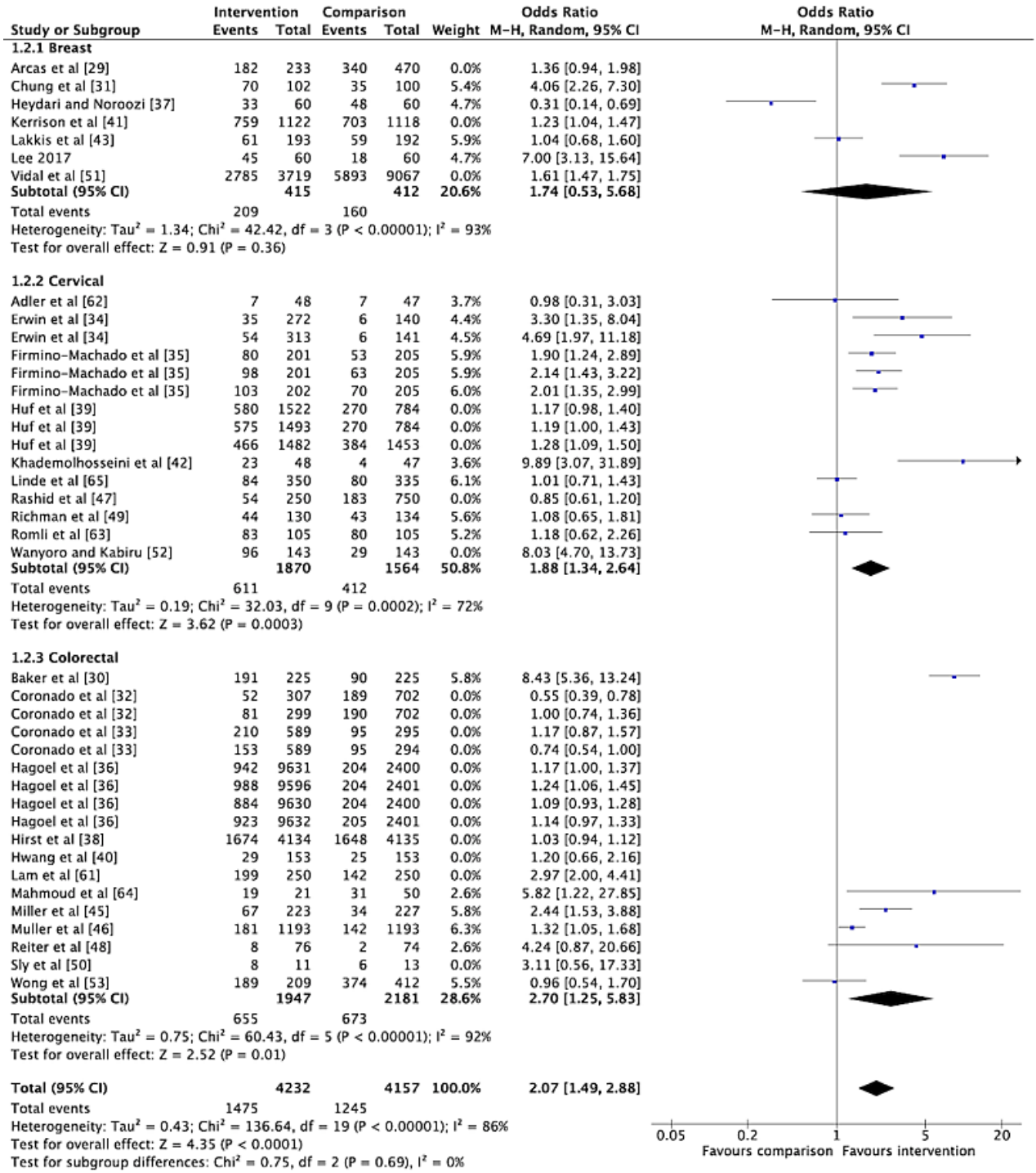


Table 3 presents the results of the secondary outcomes of screening intention. Six studies (3 RCTs and 3 pre- and postintervention studies) reported on screening intention, with two studies reporting on screening intention only. There was minor variability in the measurement of screening intention among the studies. For example, screening intention was treated as a dichotomous variable in some studies [37,45,54,58] or scored using a four-point [44] or five-point [59] Likert scale in others. Half of the studies (3/6, 50%) focused on cervical cancer, followed by breast cancer (2/6, 33%) and CRC (1/6, 17%). The intention to screen increased in all studies reporting on this outcome, except for one in which it decreased. The highest increase in screening intention was observed in the study by Lee et al [58], where there was a 24% absolute increase in the intent to receive a Papanicolaou test postintervention (19/30, 63% preintervention and 26/30, 87% postintervention). The study included a 7-day text message–based intervention that included a high level of engagement with participants through quizzes, questions, and engagement in conversation [58]. Owing to the variability in how screening intention was measured or captured, we did not perform a meta-analysis on these data.

XSL•FO
RenderX

Table 3. Cancer screening intention outcome among included studies (n=6).

Study	Study design	Outcome definition	Timeframe for assessing outcome	Outcome in comparison group (if RCT[a]) or preintervention	Outcome in intervention group (if RCT) or postintervention
Heydari and Noroozi [37]	RCT	Intention to get a mammogram (yes or no)	3 months	93% (56/60)	83% (50/60)
Lee et al [44]	RCT	Intention to receive a mammogram in the future on a 4-point scale (1=not within a year, 2=within a year, 3=within 3 months, and 4=within 1 month) among intervention and control groups	1-week postintervention	Group differences preintervention −0.64	Group differences postintervention 3.48
Miller et al [45]	RCT	Intention to receive screening measured through the postprogram iPad survey	6 months	49% (112/227)	62% (138/223)
Le and Holt [54]	Pre- and postintervention	Intent to get a Papanicolaou smear test (yes or no)	6 months	48% (22/46)	52% (24/46)
Lee et al [58]	Pre- and postintervention	Intent to receive a Papanicolaou test (yes or no)	Within 1 year	63% (19/30)	87% (26/30)
Lemos et al [59]	Pre- and postintervention	Intention to get a Papanicolaou test measured on a 5-point Likert scale from 1 (definitely will not do) to 5 (definitely will do)	6 weeks	4.50 (SD 0.64)	4.82 (SD 0.48)

[a]RCT: randomized controlled trial.

Discussion

Principal Findings

Our systematic review identified 39 studies describing the effectiveness of social media and mHealth interventions on cancer screening participation and/or intention. The overall pooled OR for cancer screening participation was significant, favoring the intervention arm (OR 1.49, 95% CI 1.31-1.70). Effect sizes were similar across all cancer types, and estimates remained stable when trials deemed to be at high risk of bias were excluded, indicating that social media, and particularly mHealth interventions, can be effective for increasing cancer screening participation.

Two systematic reviews on this topic were published in 2017 [17,18]. Uy et al [17] evaluated the effectiveness of text messaging interventions on cancer screening and identified nine studies that met the inclusion criteria. Absolute screening rates for text messaging interventions were 1%-15% higher and relative screening rates were 4%-63% higher for intervention recipients in their study [17]. The authors concluded that text messaging interventions moderately increased screening rates for breast and cervical cancer; however, additional research is needed to better quantify this relationship [17]. Tamuzi et al [18] explored mHealth interventions for cervical cancer screening only. Their review identified 17 studies, and the authors were able to perform a meta-analysis on the results by type of intervention [18]. However, their definition of mHealth was different from ours. In their study, Tamuzi et al [18] included telephone, letter, and text message reminders, whereas only text message reminders were included in our study based on our adopted definition of mHealth interventions. Text message reminders are different from these other approaches because they are sent only to mobile devices compared with telephone calls, which may be made to landlines, for which coverage has been decreasing. In addition, text messages can be sent instantly, whereas letter or postcard reminders need to be delivered by the post. Moreover, text messages have the opportunity to reach those with no fixed addresses. For example, a recent systematic review on technology use among homeless adults showed that a majority (94%) owned a cell phone [68]. Overall, Tamuzi et al [18] found that call reminders were the only intervention to show a statistically significant pooled effect estimate. Only one study included in their review reported on the effect of text message reminders, and a meta-analysis of this type of intervention was, therefore, not possible [18].

The results of this study enhance our understanding of the effectiveness of social media and mHealth interventions for cancer screening. Although both previous reviews were published in 2017, nearly 44% (17/39) of the studies in this area have been published since that time. Our review provides a comprehensive and more contemporary understanding of this topic. In addition, although previous reviews focused primarily on breast and cervical cancer, our study provides valuable insights into the effectiveness of these interventions in CRC screening as well. We included 13 studies focused on CRC in our meta-analysis and found a significant pooled effect estimate, suggesting that the use of these types of interventions can be extended to CRC as well. In comparison with the study by Uy et al [17], we found that absolute screening rates between the intervention and comparison groups were higher in our study. This may suggest that multicomponent interventions that couple social media or mHealth with additional strategies may be more

XSL•FO
RenderX

effective at increasing screening rates compared with mHealth or social media strategies alone.

The results of our study must also be understood within the larger context of interventions for cancer screening. Brouwers et al [69] conducted a systematic review of interventions for increasing cancer screening rates and looked at client reminders, client incentives, mass media, small media, group education, one-on-one education, reducing structural barriers, reducing out-of-pocket costs to clients, provider assessment and feedback, and provider incentives. Similarly, the authors found wide heterogeneity across studies and interventions and chose not to meta-analyze their data. For example, their results showed that small media interventions, including videos or printed materials such as letters, brochures, newspapers, magazines, and billboards, resulted in a point percentage increase for cancer screening participation ranging from −32.8% to 26% among studies on breast cancer, cervical cancer, and CRC [69]. Our review showed that the absolute difference between the intervention and comparison arms was 14%. The magnitude of effect varied considerably among and between intervention categories in the review by Brouwers et al [69], suggesting that additional evidence is needed for interventions related to client reminders, mass media, group education, one-on-one education, reduction of structural barriers and out-of-pocket costs, and provider incentive interventions. Given the need for additional, high-quality evidence, it is difficult to ascertain whether social media and mHealth interventions fare similar, better, or worse than non-mHealth or non–social media interventions. In addition, costs should also be considered when making any comparisons between the effectiveness of these interventions to inform the translation of these findings into practice.

Although the pooled effect estimate in our meta-analysis was consistent in the subgroup and sensitivity analyses, significant heterogeneity remained. This may be because of the variability in populations, interventions, or outcome measurement across studies. For example, the populations randomized in the studies in our review included all adults up to 79 years [44], or highly specialized populations such as emergency department patients [62] or HIV-positive individuals [60]. Moreover, many of the studies included insured samples, which may not be reflective of population-level interventions, and therefore, must be considered in the generalizability of these results. In addition, the follow-up and the intensity of each intervention varied across studies. For example, some studies may have sent a single text message reminder [37], whereas other interventions included sending multiple text messages in combination with telephone reminders [33]. Interestingly, when we looked at studies conducted in LMIC settings and excluded those with a high risk of bias, the overall pooled OR was even larger with only moderate heterogeneity. These results suggest that the effectiveness of these interventions for cancer screening participation may be more pronounced in these settings. This may be because there may be a limited number of other campaigns in these resource-low settings, whereas access to mobile phones and the internet has been reported to be comparable with that of developed nations [1].

Only a limited number of studies (n=4) tested social media interventions. As such, our results are more indicative of the effectiveness of mHealth interventions. A narrative systematic review focusing on describing the characteristics of social media interventions used for cancer prevention and management found that cancer screening participation or intention was not measured in any of the 18 studies included in the review [70]. The most common outcome measured in these studies was knowledge [70]. Although research related to social media and cancer screening participation has started to emerge [71], the inclusion of this work was limited in our review, as there are few RCTs and before and after comparisons also capturing the outcome of screening participation or intention. This suggests areas for future research to generate more evidence on the use of social media interventions for cancer screening participation. In addition, very few studies have been conducted on prostate and lung cancer screening, which is similar to what was observed in a previous study [17].

Our review and meta-analysis included a variety of mHealth and social media interventions and multicomponent interventions. Our review is comprehensive and contemporary and uses a rigorous systematic approach to screen and review the literature. As such, it includes a large number of studies for the most established screening programs for breast cancer, cervical cancer, and CRC. Owing to the large number of studies included in our review, we were able to calculate pooled effect estimates by cancer type to inform practice and future research. However, this study has limitations. Although we made every effort to obtain full-text articles, there were some records identified from our search that we could not locate. We also did not calculate a Cohen κ coefficient to report the interrater reliability between the 2 reviewers. Our review is also limited in regard to social media interventions, as only four studies were identified, with only one RCT included in the meta-analysis. This may be a reflection of current practice or due to the fact that it may be more difficult to link direct patient outcomes with the use of social media.

Conclusions

In conclusion, our results suggest that mHealth interventions may have a significant effect on cancer screening participation, particularly for breast cancer, cervical cancer, and CRC screening. Screening programs should consider the use of mHealth interventions to increase screening participation. Further research focusing on social media interventions for cancer screening participation is needed, as there was insufficient evidence available at the time of this review.

Acknowledgments

This study was supported by the Canadian Institutes of Health Research (grants FDN-148470 and GSO-157853). The funding agency had no role in the design or conduct of the study. The authors would like to acknowledge the contributions of Amina

XSL•FO
RenderX

Benmessaoud and Alawia Sherif for their help with data management and Dr Rahim Moineddin, a thesis committee member for AR's doctoral dissertation.

Conflicts of Interest

None declared.

References

1. World Health Organization. Global Difusion of eHealth: Making Universal Health Coverage Achievable. Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016:1-160.
2. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
3. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
4. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
5. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
6. Prochaska JJ, Coughlin SS, Lyons EJ. Social media and mobile technology for cancer prevention and treatment. Am Soc Clin Oncol Educ Book 2017;37:128-137 [FREE Full text] [doi: 10.1200/EDBK_173841] [Medline: 28561647]
7. Hagg E, Dahinten VS, Currie LM. The emerging use of social media for health-related purposes in low and middle-income countries: a scoping review. Int J Med Inform 2018 Jul;115:92-105. [doi: 10.1016/j.ijmedinf.2018.04.010] [Medline: 29779724]
8. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Pineros M. Global cancer observatory: cancer tomorrow. International Agency for Research on Cancer. 2018. URL: https://gco.iarc.fr/tomorrow [accessed 2020-09-15]
9. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008 Jun;103(6):1541-1549. [doi: 10.1111/j.1572-0241.2008.01875.x] [Medline: 18479499]
10. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L, U.S. Preventive Services Task Force. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med 2009 Nov 17;151(10):727-737 [FREE Full text] [doi: 10.7326/0003-4819-151-10-200911170-00009] [Medline: 19920273]
11. Peirson L, Fitzpatrick-Lewis D, Ciliska D, Warren R. Screening for cervical cancer: a systematic review and meta-analysis. Syst Rev 2013 May 24;2:35 [FREE Full text] [doi: 10.1186/2046-4053-2-35] [Medline: 23706117]
12. Sadate A, Occean BV, Beregi J, Hamard A, Addala T, de Forges H, et al. Systematic review and meta-analysis on the impact of lung cancer screening by low-dose computed tomography. Eur J Cancer 2020 Jul;134:107-114. [doi: 10.1016/j.ejca.2020.04.035] [Medline: 32502939]
13. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut 2015 Oct;64(10):1637-1649. [doi: 10.1136/gutjnl-2014-309086] [Medline: 26041752]
14. Cancer Screening in Canada: An overview of screening participation for breast, cervical, and colorectal cancer. Toronto: Canadian Partnership Against Cancer. 2015. URL: https://s22457.pcdn.co/wp-content/uploads/2019/01/Breast-Cervical-Colorectal-Screening-Participate-2015-EN.pdf [accessed 2020-09-15]
15. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol 2012 Jun;36(3):237-248. [doi: 10.1016/j.canep.2012.02.007] [Medline: 22459198]
16. Singh H, Bernstein CN, Samadder JN, Ahmed R. Screening rates for colorectal cancer in Canada: a cross-sectional study. CMAJ Open 2015;3(2):149-157 [FREE Full text] [doi: 10.9778/cmajo.20140073] [Medline: 26389092]
17. Uy C, Lopez J, Trinh-Shevrin C, Kwon SC, Sherman SE, Liang PS. Text messaging interventions on cancer screening rates: a systematic review. J Med Internet Res 2017 Aug 24;19(8):e296 [FREE Full text] [doi: 10.2196/jmir.7893] [Medline: 28838885]
18. Tamuzi JL. Effectiveness of mHealth to increase cervical cancer screening: systematic review of interventions. Int J Pul Res Sci 2017 Oct 27;2(3):555586. [doi: 10.19080/ijoprs.2017.02.555586]
19. Parackal M, Parackal S, Eusebius S, Mather D. The use of Facebook advertising for communicating public health messages: a campaign against drinking during pregnancy in New Zealand. JMIR Public Health Surveill 2017 Aug 10;3(3):e49 [FREE Full text] [doi: 10.2196/publichealth.7032] [Medline: 28798011]

XSL•FO
RenderX

20. Cavallo DN, Chou WS, McQueen A, Ramirez A, Riley WT. Cancer prevention and control interventions using social media: user-generated approaches. Cancer Epidemiol Biomarkers Prev 2014 Sep;23(9):1953-1956. [doi: 10.1158/1055-9965.EPI-14-0593] [Medline: 25103820]
21. Valle CG, Tate DF. Engagement of young adult cancer survivors within a Facebook-based physical activity intervention. Transl Behav Med 2017 Dec;7(4):667-679 [FREE Full text] [doi: 10.1007/s13142-017-0483-3] [Medline: 28374211]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
23. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. Press peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Dec;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
24. Ruco A, Dossa F, Tinmouth J, Llovet D, Kishibe T, Baxter NN. Social media and mobile health technology for cancer screening: a systematic review and meta-analysis protocol. BMJ Open 2020 Feb 05;10(2):e035411 [FREE Full text] [doi: 10.1136/bmjopen-2019-035411] [Medline: 32029500]
25. Sterne JA, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
26. EPOC Resources for review authors. Cochrane Effective Practice and Organisation of Care (EPOC). 2017. URL: https://tinyurl.com/2du6ffc4 [accessed 2020-09-15]
27. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2020 Apr 26:55-61 [FREE Full text] [doi: 10.1002/jrsm.1411] [Medline: 32336025]
28. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. Chichester, UK: John Wiley & Sons; 2019:1-728.
29. Arcas MM, Buron A, Ramis O, Esturi M, Hernández C, Macià F. [Can a mobile phone short message increase participation in breast cancer screening programmes?]. Rev Calid Asist 2014;29(4):188-196. [doi: 10.1016/j.cali.2014.02.003] [Medline: 25002239]
30. Baker DW, Brown T, Buchanan DR, Weil J, Balsley K, Ranalli L, et al. Comparative effectiveness of a multifaceted intervention to improve adherence to annual colorectal cancer screening in community health centers: a randomized clinical trial. JAMA Intern Med 2014 Aug;174(8):1235-1241. [doi: 10.1001/jamainternmed.2014.2352] [Medline: 24934845]
31. Chung IY, Kang E, Yom CK, Kim D, Sun Y, Hwang Y, et al. Effect of short message service as a reminder on breast self-examination in breast cancer patients: a randomized controlled trial. J Telemed Telecare 2015 Apr;21(3):144-150. [doi: 10.1177/1357633X15571651] [Medline: 25697492]
32. Coronado GD, Rivelli JS, Fuoco MJ, Vollmer WM, Petrik AF, Keast E, et al. Effect of reminding patients to complete fecal immunochemical testing: a comparative effectiveness study of automated and live approaches. J Gen Intern Med 2018 Jan;33(1):72-78 [FREE Full text] [doi: 10.1007/s11606-017-4184-x] [Medline: 29019046]
33. Coronado GD, Thompson JH, Petrik AF, Nyongesa DB, Leo MC, Castillo M, et al. Patient-refined messaging for a mailed colorectal cancer screening program: findings from the PROMPT study. J Am Board Fam Med 2019;32(3):318-328 [FREE Full text] [doi: 10.3122/jabfm.2019.03.180275] [Medline: 31068396]
34. Erwin E, Aronson KJ, Day A, Ginsburg O, Macheku G, Feksi A, et al. SMS behaviour change communication and eVoucher interventions to increase uptake of cervical cancer screening in the Kilimanjaro and Arusha regions of Tanzania: a randomised, double-blind, controlled trial of effectiveness. BMJ Innov 2019 Jan;5(1):28-34 [FREE Full text] [doi: 10.1136/bmjinnov-2018-000276] [Medline: 31645991]
35. Firmino-Machado J, Varela S, Mendes R, Moreira A, Lunet N, SCAN-Cervical Cancer collaborators. A 3-step intervention to improve adherence to cervical cancer screening: the SCAN randomized controlled trial. Prev Med 2019 Jun;123:250-261. [doi: 10.1016/j.ypmed.2019.03.025] [Medline: 30936001]
36. Hagoel L, Neter E, Stein N, Rennert G. Harnessing the question-behavior effect to enhance colorectal cancer screening in an mHealth experiment. Am J Public Health 2016 Nov;106(11):1998-2004. [doi: 10.2105/AJPH.2016.303364] [Medline: 27631750]
37. Heydari E, Noroozi A. Comparison of two different educational methods for teachers' mammography based on the health belief model. Asian Pac J Cancer Prev 2015;16(16):6981-6986 [FREE Full text] [doi: 10.7314/apjcp.2015.16.16.6981] [Medline: 26514478]
38. Hirst Y, Skrobanski H, Kerrison RS, Kobayashi LC, Counsell N, Djedovic N, et al. Text-message Reminders in Colorectal Cancer Screening (TRICCS): a randomised controlled trial. Br J Cancer 2017 May 23;116(11):1408-1414 [FREE Full text] [doi: 10.1038/bjc.2017.117] [Medline: 28441381]
39. Huf S, Kerrison RS, King D, Chadborn T, Richmond A, Cunningham D, et al. Behavioral economics informed message content in text message reminders to improve cervical screening participation: two pragmatic randomized controlled trials. Prev Med 2020 Oct;139:106170. [doi: 10.1016/j.ypmed.2020.106170] [Medline: 32610059]

XSL•FO
RenderX

40. Hwang KO, Ottenbacher AJ, Graham AL, Thomas EJ, Street RL, Vernon SW. Online narratives and peer support for colorectal cancer screening: a pilot randomized trial. Am J Prev Med 2013 Jul;45(1):98-107. [doi: 10.1016/j.amepre.2013.02.024] [Medline: 23790994]
41. Kerrison RS, Shukla H, Cunningham D, Oyebode O, Friedman E. Text-message reminders increase uptake of routine breast screening appointments: a randomised controlled trial in a hard-to-reach population. Br J Cancer 2015 Mar 17;112(6):1005-1010 [FREE Full text] [doi: 10.1038/bjc.2015.36] [Medline: 25668008]
42. Khademolhosseini F, Noroozi A, Tahmasebi R. The effect of health belief model-based education through Telegram instant messaging services on Pap smear performance. Asian Pac J Cancer Prev 2017 Aug 27;18(8):2221-2226 [FREE Full text] [doi: 10.22034/APJCP.2017.18.8.2221] [Medline: 28843259]
43. Lakkis NA, Atfeh AM, El-Zein YR, Mahmassani DM, Hamadeh GN. The effect of two types of sms-texts on the uptake of screening mammogram: a randomized controlled trial. Prev Med 2011 Oct;53(4-5):325-327. [doi: 10.1016/j.ypmed.2011.08.013] [Medline: 21871480]
44. Lee H, Ghebre R, Le C, Jang YJ, Sharratt M, Yee D. Mobile phone multilevel and multimedia messaging intervention for breast cancer screening: pilot randomized controlled trial. JMIR Mhealth Uhealth 2017 Nov 07;5(11):e154 [FREE Full text] [doi: 10.2196/mhealth.7091] [Medline: 29113961]
45. Miller DP, Denizard-Thompson N, Weaver KE, Case LD, Troyer JL, Spangler JG, et al. Effect of a digital health intervention on receipt of colorectal cancer screening in vulnerable patients: a randomized controlled trial. Ann Intern Med 2018 Apr 17;168(8):550-557 [FREE Full text] [doi: 10.7326/M17-2315] [Medline: 29532054]
46. Muller CJ, Robinson RF, Smith JJ, Jernigan MA, Hiratsuka V, Dillard DA, et al. Text message reminders increased colorectal cancer screening in a randomized trial with Alaska Native and American Indian people. Cancer 2017 Apr 15;123(8):1382-1389. [doi: 10.1002/cncr.30499] [Medline: 28001304]
47. Rashid RM, Mohamed M, Hamid ZA, Dahlui M. Is the phone call the most effective method for recall in cervical cancer screening?--results from a randomised control trial. Asian Pac J Cancer Prev 2013;14(10):5901-5904 [FREE Full text] [doi: 10.7314/apjcp.2013.14.10.5901] [Medline: 24289597]
48. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing human papillomavirus vaccination among young gay and bisexual men: a randomized pilot trial of the outsmart HPV intervention. LGBT Health 2018 Jul;5(5):325-329 [FREE Full text] [doi: 10.1089/lgbt.2018.0059] [Medline: 29979642]
49. Richman AR, Maddy L, Torres E, Goldberg EJ. A randomized intervention study to evaluate whether electronic messaging can increase human papillomavirus vaccine completion and knowledge among college students. J Am Coll Health 2016;64(4):269-278. [doi: 10.1080/07448481.2015.1117466] [Medline: 26821923]
50. Sly JR, Miller SJ, Jandorf L. The digital divide and health disparities: a pilot study examining the use of short message service (SMS) for colonoscopy reminders. J Rac Ethn Health Dispar 2014 Jul 11;1(4):231-237. [doi: 10.1007/s40615-014-0029-z]
51. Vidal C, Garcia M, Benito L, Milà N, Binefa G, Moreno V. Use of text-message reminders to improve participation in a population-based breast cancer screening program. J Med Syst 2014 Sep;38(9):118-124. [doi: 10.1007/s10916-014-0118-x] [Medline: 25073694]
52. Wanyoro A, Kabiru E. Use of mobile phone short text message service to enhance cervical cancer screening at Thika Level 5 hospital, Kiambu County, Kenya: a randomised controlled trial. Res Obstet Gynaecol 2017;5(1):10-20. [doi: 10.5923/j.rog.20170501.03]
53. Wong MC, Ching JY, Huang J, Wong JC, Lam TY, Chan VC, et al. Effectiveness of reminder strategies on cancer screening adherence: a randomised controlled trial. Br J Gen Pract 2018 Sep;68(674):604-611 [FREE Full text] [doi: 10.3399/bjgp18X698369] [Medline: 30104327]
54. Le D, Holt CL. CervixCheck: A spiritually-based text messaging intervention to promote cervical cancer awareness and Pap test screening intention among African-American women. J Health Commun 2018;23(9):842-853. [doi: 10.1080/10810730.2018.1528317] [Medline: 30300091]
55. Capık C, Gözüm S. The effect of web-assisted education and reminders on health belief, level of knowledge and early diagnosis behaviors regarding prostate cancer screening. Eur J Oncol Nurs 2012 Feb;16(1):71-77 [FREE Full text] [doi: 10.1016/j.ejon.2011.03.007] [Medline: 21530397]
56. Jessup DL, Glover IM, Daye D, Banzi L, Jones P, Choy G, et al. Implementation of digital awareness strategies to engage patients and providers in a lung cancer screening program: retrospective study. J Med Internet Res 2018 Feb 15;20(2):e52 [FREE Full text] [doi: 10.2196/jmir.8932] [Medline: 29449199]
57. Fornos LB, Urbansky KA, Villarreal R. Increasing cervical cancer screening for a multiethnic population of women in South Texas. J Cancer Educ 2014 Mar;29(1):62-68. [doi: 10.1007/s13187-013-0544-3] [Medline: 24170274]
58. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
59. Lemos M, Rothes I, Oliveira F, Soares L. Raising cervical cancer awareness: analysing the incremental efficacy of Short Message Service. Health Edu J 2017 Sep 14;76(8):956-970 [FREE Full text] [doi: 10.1177/0017896917728306]

XSL•FO
RenderX

60. Ganta V, Moonie S, Patel D, Hunt AT, Richardson J, Di John D, et al. Timely reminder interventions to improve annual Papanicolaou (Pap) smear rates among HIV-infected women in an outpatient center of southern Nevada: a short report. AIDS Care 2017 Dec;29(9):1099-1101. [doi: 10.1080/09540121.2017.1322677] [Medline: 28460538]
61. Lam TY, Wong MC, Ching JY, Chan V, Ng SK, Hui SN, et al. 210 - Effectiveness of Whatsapp reminder on compliance with colorectal cancer screening: a randomized controlled trial. Gastroenterology 2018 May;154(6):58-59. [doi: 10.1016/s0016-5085(18)30653-x]
62. Adler D, Abar B, Wood N, Bonham A. An intervention to increase uptake of cervical cancer screening among emergency department patients: results of a randomized pilot study. J Emerg Med 2019 Dec;57(6):836-843 [FREE Full text] [doi: 10.1016/j.jemermed.2019.07.021] [Medline: 31594738]
63. Romli R, Shahabudin S, Saddki N, Mokhtar N. Effectiveness of a health education program to improve knowledge and attitude towards cervical cancer and Pap smear: a controlled community trial in Malaysia. Asian Pac J Cancer Prev 2020 Mar 01;21(3):853-859 [FREE Full text] [doi: 10.31557/APJCP.2020.21.3.853] [Medline: 32212817]
64. Mahmud N, Doshi SD, Coniglio MS, Clermont M, Bernard D, Reitz C, et al. An automated text message navigation program improves the show rate for outpatient colonoscopy. Health Educ Behav 2019 Dec;46(6):942-946 [FREE Full text] [doi: 10.1177/1090198119869964] [Medline: 31431077]
65. Linde DS, Andersen MS, Mwaiselage J, Manongi R, Kjaer SK, Rasch V. Effectiveness of one-way text messaging on attendance to follow-up cervical cancer screening among human papillomavirus-positive Tanzanian women (Connected2Care): parallel-group randomized controlled trial. J Med Internet Res 2020 Apr 02;22(4):e15863 [FREE Full text] [doi: 10.2196/15863] [Medline: 32238335]
66. Lyson HC, Le GM, Zhang J, Rivadeneira N, Lyles C, Radcliffe K, et al. Social media as a tool to promote health awareness: results from an online cervical cancer prevention study. J Cancer Educ 2019 Aug;34(4):819-822. [doi: 10.1007/s13187-018-1379-8] [Medline: 29948924]
67. Key KV, Adegboyega A, Bush H, Aleshire ME, Contreras OA, Hatcher J. #CRCFREE: Using social media to reduce colorectal cancer risk in rural adults. Am J Health Behav 2020 May 01;44(3):353-363. [doi: 10.5993/AJHB.44.3.8] [Medline: 32295683]
68. Rhoades H, Wenzel SL, Rice E, Winetrobe H, Henwood B. No digital divide? Technology use among homeless adults. J Soc Distress Homeless 2017 Mar 22;26(1):73-77 [FREE Full text] [doi: 10.1080/10530789.2017.1305140] [Medline: 31097900]
69. Brouwers MC, De Vito C, Bahirathan L, Carol A, Carroll JC, Cotterchio M, et al. What implementation interventions increase cancer screening rates? A systematic review. Implement Sci 2011 Sep 29;6:111 [FREE Full text] [doi: 10.1186/1748-5908-6-111] [Medline: 21958556]
70. Han CJ, Lee YJ, Demiris G. Interventions using social media for cancer prevention and management: a systematic review. Cancer Nurs 2018;41(6):19-31 [FREE Full text] [doi: 10.1097/NCC.0000000000000534] [Medline: 28753192]
71. Koïvogui A, Levi S, Finkler M, Lewkowicz S, Gombeaud T, Sabate JM, et al. Feasibility of encouraging participation in colorectal cancer screening campaigns by motivating people through the social network, Facebook. Colorectal Dis 2020 Oct;22(10):1325-1335. [doi: 10.1111/codi.15121] [Medline: 32397003]

Abbreviations

CRC: colorectal cancer
LMIC: low- and middle-income country
MeSH: Medical Subject Headings
mHealth: mobile health
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by R Kukafka; submitted 31.12.20; peer-reviewed by L Guo, E Neter; comments to author 08.02.21; revised version received 15.02.21; accepted 21.06.21; published 30.07.21.

Please cite as:
Ruco A, Dossa F, Tinmouth J, Llovet D, Jacobson J, Kishibe T, Baxter N
Social Media and mHealth Technology for Cancer Screening: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(7):e26759
URL: https://www.jmir.org/2021/7/e26759
doi:10.2196/26759
PMID:34328423

XSL•FO
RenderX

©Arlinda Ruco, Fahima Dossa, Jill Tinmouth, Diego Llovet, Jenna Jacobson, Teruko Kishibe, Nancy Baxter. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

The Ethical Digital Surgeon

Kyle Lam[1], MB BChir; Sanjay Purkayastha[1], MD; James M Kinross[1], PhD

Department of Surgery and Cancer, Imperial College London, London, United Kingdom

Corresponding Author:
Kyle Lam, MB BChir
Department of Surgery and Cancer
Imperial College London
10th Floor QEQM Building, Praed St
St Mary's Hospital
London, W2 1NY
United Kingdom
Phone: 44 07964904213
Email: k.lam@imperial.ac.uk

Abstract

This viewpoint explores the ethical and regulatory consequences of the digital transformation of the operating room. Surgical robotics is undergoing significant change and future advances will center around the capture and use of data. The consequences of creating this surgical data pipeline must be understood and digital surgical systems must prioritize the safeguarding of patient data. Moreover, data protection laws and frameworks must adapt to the changing nature of surgical data. Finally, digital surgeons must understand changing data legislation and best practice on data governance to act as guardians not only for their own but also for their patients' data.

(J Med Internet Res 2021;23(7):e25849) doi:10.2196/25849

KEYWORDS

digital surgery; ethics; data governance; robotics; digital surgeons; surgery; digital health care; smartphone app

Digital Surgery: Overview

Robotic surgery is currently undergoing significant change. Global medical device companies are entering a rapidly growing market based on economic optimism and a shifting clinical evidence base. COVID-19 has dramatically sped up this transformation with rapid global uptake in numerous digital platforms across health care [1]. The majority of these robots are systems built on incremental mechanical innovations that require a surgeon to operate a console that drives single, multiport, or endoscopic robotic arms. However, the real revolution is happening in the data that will be captured from sensors and systems embedded within these robotic platforms and streamed to the cloud. "Digital surgery" is defined as the capture, storage, analysis, and visualization of surgical data sets for precision surgery. It incorporates numerous enabling technologies such as artificial intelligence (AI), cloud computing and image augmentation, such as the use of 3D augmented reality guidance for the removal of complex tumors [2]. The hope is that digital robotic systems can be leveraged for real-time decision support surgery and it is this that will finally deliver the quantum leap in safety and clinical outcomes that robotic surgery has to date not yet provided. This compelling vision has created significant hype and general surgery robotic adoption rates are rapidly climbing [3-5].

Ethical and Regulatory Challenges of Digital Surgery

What is far less clear are the ethical and regulatory consequences of suddenly turning on a limitless surgical data pipeline. Operating room data sets are unique, large, and heterogenous (eg, imaging, video, sensor data, text, team, and instrument performance) and each contains a priceless data surplus that will be mined for as yet unimagined purposes. It is unclear how patients undergoing surgery in the future operating room will be protected from data misuse or how hospital providers and patients will be reimbursed by the companies that use them to create their digital products.

However, these innovations do not just create ethical challenges for the patient. The future surgeon will also operate in a world where every minor action and decision will be subject to scrutiny. This moves beyond the capture of relatively blunt operative performance measures such as time, conversion rates, patient outcomes, and theater utilization into a paradigm of machine learning of an individuals' total surgical performance.

XSL•FO
RenderX

"Digital surgeons" who embrace this world will have much to gain through faster and safer learning curves, continuous education, and the delivery of safer procedures. Although digital surgeons will probably remain accountable for the decisions that they make, it is now possible that they will also have to contend with automation bias, opaque algorithms, and a rapidly evolving ecosystem of sophisticated cloud-based platforms and connected hardware. This will in turn create new challenges for consent and litigation, which are as of yet untested. The digital surgeon will serve as a guardian not only for their own data, but also for their patient's data. It is therefore imperative that they understand and keep up to date with data legislation and best practice on data governance. "Binary surgeons" who do not have access to digital robotic platforms or who choose to reject them will be left in a perilous position where their performance will be compared with digitally augmented colleagues regardless of whether they "opt out."

"Streams"—Lessons Learned?

There is now a critical need for the surgical community to address the ethical and regulatory challenges of this brave new world. We can heed many lessons from the introduction of similar technologies in the public clinical domain. In July 2015, clinicians from the Royal Free London NHS Foundation Trust approached Google DeepMind Technologies Limited in order to create a smartphone app called "Streams" to aid clinicians in the management of acute kidney injury. The company was given access to the complete health records of all 1.6 million patients attending Barnet, Chase Farm, and Royal Free Hospitals over more than a 5-year period. Neither DeepMind nor the Royal Free obtained the requisite approval from the Information Commissioner's Office, the Health Research Authority, or the Confidentiality Advisory Group [6]. In 2017, the Information Commissioner's Office found several shortcomings in how data were handled, ordering an independent audit into "Streams" to be conducted [7]. While the audit concluded that the Royal Free and DeepMind's actions were not unlawful [8], the episode demonstrates that patients should be aware of when and why their data are being used and have the ability to opt out if desired. The UK's National Data Opt-Out, introduced in May 2018 [9], which allows patients to opt out of their confidential patient information being used for research and planning, is a key step in the right direction in this regard.

Future Challenges

However, this may be more challenging in an operative environment. For example, if digital surgery systems are inherently dependent on data capture and analysis to function through networked robotic systems, what are the consequences for patients or surgeons who refuse or do not wish to share their data? National bodies such as the UK's NHSX, responsible for the digitization of the National Health Service, have already established their national center of expertise overseeing data sharing agreements with industry [10]. The hope is that these organizations can create frameworks for delivering standardized guidance and contracts to support successful relationships between health care and industry in an equitable manner.

Irrespective of these initiatives, both patients and surgeons must have total oversight of what digital surgery companies are doing with our data and future algorithm development. US and EU regulators are currently trying to ensure that artificial intelligent systems are safely introduced into the operating room. Data protection laws such as the EU's General Data Protection Regulation (GDPR) and the USA's Health Insurance Portability and Accountability Act (HIPAA) also provide a framework for data protection. It is questionable whether these frameworks do enough to protect surgical patients; the "surgical data surplus" is priceless and health care organizations must guard this revenue source closely if we are to build sustainable digital surgical systems and keep control of the health inequality gap.

Studies have shown that health care providers bear the weight of public expectation when patients consent to share their data [11]. Digital surgeons must therefore have basic AI literacy and foundational knowledge in data governance and protection. Moreover, the digital surgeon must now lead in the creation of "trustworthy AI" which is legal, ethical, and fit for surgical purpose. In turn, future robotic systems must be developed, tested, and trialed in a learning system that incorporates digital surgery as a core pillar of its development. Finally, international surgical bodies must also determine what role they are going to play in this brave new world and how they are going to certify and train the digital surgeon of the future (Textbox 1).

Textbox 1. Key messages.

- Digital surgical systems must prioritize the safeguarding of patient data.
- Data protection laws and frameworks must adapt to the changing nature of surgical data.
- Future surgeons must be artificial intelligence literate and understand data governance and protection.

Conflicts of Interest

SP provides consultations for Medtronic, T.M.L.E. Ltd., and Roche. SP is also the cofounder and director of Mangetoo, 1 World Medical, and the London General Surgery Clinic. SP is also a partner of OneWelbeck Hospital. JK provides consultations for Verb Robotics, SafeHeal, YSOPIA Bioscience, and Universal Diagnostics (UDX). JK also received equity from Mangetoo (teledietetics), OneWelbeck Day Surgery (Hospital), 1 World Medical (personal protective equipment), and Medical iSight (augmented reality). KL has no conflicts of interest to declare for this paper.

XSL•FO
RenderX

References

1. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. The Lancet Digital Health 2020 Aug;2(8):e435-e440. [doi: 10.1016/s2589-7500(20)30142-4]
2. Porpiglia F, Checcucci E, Amparore D, Piramide F, Volpi G, Granato S, et al. Three-dimensional Augmented Reality Robot-assisted Partial Nephrectomy in Case of Complex Tumours (PADUA ≥10): A New Intraoperative Tool Overcoming the Ultrasound Guidance. European Urology 2020 Aug;78(2):229-238. [doi: 10.1016/j.eururo.2019.11.024]
3. Lam K, Clarke J, Purkayastha S, Kinross JM. Uptake and accessibility of surgical robotics in England. Int J Med Robot 2020 Oct 07;17(1):1-7. [doi: 10.1002/rcs.2174]
4. Sheetz KH, Claflin J, Dimick JB. Trends in the Adoption of Robotic Surgery for Common Surgical Procedures. JAMA Netw Open 2020 Jan 10;3(1):e1918911. [doi: 10.1001/jamanetworkopen.2019.18911]
5. Intuitive. Annual Report. 2019. URL: https://isrg.gcs-web.com/static-files/31b5c428-1d95-4c01-9c85-a7293bac5e05 [accessed 2021-01-13]
6. Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol 2017 Mar 16;7(4):351-367. [doi: 10.1007/s12553-017-0179-1]
7. Information Commissioner's Office. Royal Free - Google DeepMind trial failed to comply with data protection law. 2017. URL: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/ [accessed 2021-01-13]
8. Linklaters. Audit of the acute kidney injury detection system known as Streams. URL: http://s3-eu-west-1.amazonaws.com/files.royalfree.nhs.uk/Reporting/Streams_Report.pdf [accessed 2021-01-13]
9. NHS Digital. National data opt-out. URL: https://digital.nhs.uk/services/national-data-opt-out [accessed 2021-01-13]
10. Department of Health and Social Care. Creating the right framework to realise the benefits for patients and the NHS where data underpins innovation. 2019. URL: https://www.gov.uk/government/publications/creating-the-right-framework-to-realise-the-benefits-of-health-data/creating-the-right-framework-to-realise-the-benefits-for-patients-and-the-nhs-where-data-underpins-innovation [accessed 2021-01-13]
11. Ipsos MORI Social Research Institute. The One-Way Mirror: Public attitudes to commercial access to health data. 2016. URL: https://wellcome.org/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf [accessed 2021-01-13]

Abbreviations

AI: artificial intelligence
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act

Edited by R Kukafka; submitted 18.11.20; peer-reviewed by F Iqbal, S Pit; comments to author 13.01.21; revised version received 19.01.21; accepted 30.04.21; published 05.07.21.

Please cite as:
Lam K, Purkayastha S, Kinross JM
The Ethical Digital Surgeon
J Med Internet Res 2021;23(7):e25849
URL: https://www.jmir.org/2021/7/e25849
doi:10.2196/25849
PMID:34453502

©Kyle Lam, Sanjay Purkayastha, James M Kinross. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Overcoming the Digital Divide in the Post–COVID-19 "Reset": Enhancing Group Virtual Visits with Community Health Workers

Megha K Shah[1*], MD, MSc; Ashley Christina Gibbs[1*], MSc; Mohammed K Ali[1,2*], MD, MSc, MBA; K M Venkat Narayan[2*], MD, MSc, MBA; Nadia Islam[3*], PhD

[1]Department of Family and Preventive Medicine, School of Medicine, Emory University, Dunwoody, GA, United States

[2]Hubert Department of Global Health, Emory Rollins School of Public Health, Emory University, Atlanta, GA, United States

[3]Division of Population Health, Grossman School of Medicine, New York University, New York City, NY, United States

[*]all authors contributed equally

Corresponding Author:
Megha K Shah, MD, MSc
Department of Family and Preventive Medicine
School of Medicine
Emory University
4500 N Shallowford Rd
Dunwoody, GA, 30338
United States
Phone: 1 4047786944
Fax: 1 4047786901
Email: mkshah@emory.edu

Abstract

The COVID-19 pandemic created numerous barriers to the implementation of participant-facing research. For most, the pandemic required rapid transitioning to all virtual platforms. During this pandemic, the most vulnerable populations are at highest risk of falling through the cracks of engagement in clinical care and research. Nonetheless, we argue that we should reframe the discussion to consider how this transition may create opportunities to engage extensively to reach populations. Here, we present our experience in Atlanta (Georgia, United States) in transitioning a group visit model for South Asian immigrants to a virtual platform and the pivotal role community members in the form of community health workers can play in building capacity among participants. We provide details on how this model helped address common barriers to group visit models in clinical practice and how our community health worker team innovatively addressed the digital challenges of working with an elderly population with limited English proficiency.

(J Med Internet Res 2021;23(7):e27682) doi:10.2196/27682

KEYWORDS

community health workers; COVID-19; diabetes mellitus; eHealth; elderly; health equity; telemedicine; virtual; vulnerable populations

Introduction

The rapid adoption and escalation of telemedicine during the current COVID-19 pandemic has led to concerns about widening gaps in health equity [1]. Among populations with limited digital access or literacy, notable gaps in access to care have been found; for example, less than one-third of Medicare beneficiaries aged over 65 years have reported digital access at home, and those aged over 75 years and with less than high school–level education are less likely to use technology for health care needs [2,3]. Potentially vulnerable groups include underserved racial and ethnic groups, older adults, and adults with limited English proficiency (LEP), for whom barriers to care access and health information technology have been well-documented, exacerbated by the pandemic, and are at risk of exacerbation [3-8].

In addition, the pandemic and social distancing guidelines have profoundly limited the ability of vulnerable populations to access reliable health care [9,10]. Particularly for chronic disease prevention and management, this can lead to delays in care, which may result in poor health outcomes. Over 40% of respondents in a recent survey of over 5000 adults in the United States reported delaying or avoiding care owing to concerns related to COVID-19, which has been linked to excess deaths

XSL•FO
RenderX

reported in 2020, compared to prior years [9]. Older adults, people of color, and low-income individuals—the same populations with the lowest telemedicine access and literacy—have been disproportionately afflicted by COVID-19 and are at the highest risk of cardiometabolic disease [11,12]. The confluence of health risk and the digital divide in these communities creates an environment in which these groups may be most vulnerable to gaps in care and social isolation, which are factors known to exacerbate chronic health conditions.

Despite these challenges, virtual group visits led by community health workers (CHWs)—trained public health professionals with shared lived experience of the communities they serve—may be a pivotal strategy to help address the digital divide in telehealth and foster social connectedness. Social connections can not only influence the risk of chronic illness but also improve chronic disease management [13,14]. Social support is considered a major component of chronic disease self-management [15]. Previous studies suggest that adults with limited social interactions are less likely to engage in health behaviors such as physical activity, smoking cessation, and healthy eating habits [16]. To address this and stimulate socially cohesive and supportive care, group visits have emerged as a type of visit format in which a portion of the visit is a group education class, which can be led by a health coach, nutritionist, or medical assistant, and a portion of time is spent in a brief one-on-one visit with a clinician. Nonetheless, the implementation of group visits has some challenges [17,18]. These challenges include visit-level challenges such as the logistics of clinic workflow and meeting space, patient-level factors such as scheduling or transportation conflicts, and group-leader level challenges such as comfort level in a group role and engaging and retaining participants.

Group visits can be enhanced by meeting patients where they live, work, and worship. CHWs are trained and trusted members of their community, who can serve as a bridge between clinical and community settings. CHWs can work in a variety of capacities to support health education, treatment adherence, health system navigation, and linkage to social services for some of the most vulnerable populations with chronic health conditions [19]. Integrating CHWs into and having them lead group visits may be a particularly effective approach to engaging patients in socially cohesive activities because of CHWs' familiarity with the communities that they work with. Often, they have shared experiences that enhance rapport-building with patients. A systematic review reported that chronic disease management interventions that included CHWs were associated with increases in tobacco cessation, improved blood pressure control, and low blood sugar levels, with no risk of adverse events [20].

In the current context of the COVID-19 pandemic, virtual group visits may have additional challenges but also expose unanticipated advantages to address the digital divide. Here, we describe our experience transitioning from in-person to a virtual CHW-guided group visit intervention for diabetes and hypertension management among older South Asian adults with LEP.

DREAM Atlanta

Funded by the National Institutes of Health, the DREAM Atlanta study was designed to test the effectiveness of CHW-delivered in-person health education on diabetes and hypertension, which was tailored for South Asian adults living in Atlanta. Over the course of 6 months, those in the intervention group were intended to receive 5 group education sessions and 2 one-on-one home visits. Initiated in the fall of 2019, the pandemic forced the program's transition to virtual group visits immediately at the start of recruitment. The transition to a virtual program necessitated not only training of CHWs and other study staff in utilizing technology to deliver remote sessions, but also to educate study participants to utilize digital devices and access virtual platforms. Among the 190 adults of South Asian descent enrolled in the program, all reported English as a second language. At baseline, participants were on average, 56 (range 30-80) years old, approximately 40% of whom were over the age of 60 years; 56% were female; 44% were male; and 96% reported having access to a smartphone or tablet device.

The CHWs worked one-on-one with participants to address barriers to engaging with remote technologies such as creating email accounts, downloading apps on smart devices, and teaching participants how to use features of videoconferencing such as the video and mute features. CHWs utilized several strategies to help participants connect.

Rapid Transition Assessment Methods and Findings

To gather the perspectives of the study team on the transition to virtual group visits, a 90-minute, tape-recorded video group discussion was conducted with 3 CHWS and the study coordinator. Discussion topics included a discussion of feasibility to conduct virtual recruitment and study sessions, barriers to implementing the study virtually, and specific adaption processes utilized. These data were supplemented with reviews of meeting minutes from weekly team meetings held from March to December 2020. The first and senior authors reviewed common themes emerging from the group discussion and reviews of meeting minutes to determine key facilitators and barriers to the implementation of the remote intervention. All study activities conducted were reviewed by the institutional review board of the New York University School of Medicine, which served as the single institutional review board for this study.

Table 1 lists the digital challenges and solutions fostered by the CHW team. For example, CHWs would start with technologies that participants were familiar with to help them learn to log on to the videoconferencing apps. Most of the DREAM study participants accessed the internet on their smartphones, and similar to studies in South Asia, participants report familiarity with platforms such as Facebook and WhatsApp [21]. Thus, for example, if participants used WhatsApp, the CHW would ask them to join a video call on this platform to help them understand the process of connecting to study-approved remote platforms such as Zoom. This teach back method by the CHWs

XSL•FO
RenderX

to their participants greatly enhanced their confidence to schedule and lead group virtual visit sessions.

In the 5 months of this transition, we have observed CHWs' unique strengths to help close the digital divide in this highly vulnerable population. While previous studies have demonstrated the effectiveness of mobile technologies for chronic health conditions, including diabetes self-management, few studies have examined or reported on the process of getting participants connected to telehealth platforms [22,23]. However, technical challenges and education of patients about telehealth services are well-documented challenges for health systems' implementation of telehealth services [24]. Similar to our experience, a recent systematic review identified participant training in videoconferencing by an information technology specialist or a group facilitator as a method to overcome participant challenges to connect for psychotherapy interventions [25]. Thus, our findings add to and may have broader implications to address the known challenges of user-related technical difficulties associated with virtual health services.

Table 1. Digital challenges and strategies used by community health workers to enhance attendance in virtual group visits among South Asian adults with limited English proficiency.

Digital challenge	Strategy or modification	Example or representative quote
Community health workers had limited experience with videoconferencing software	• In-person sessions with the project coordinator and trial and error with the platform to learn to use features	• "The project coordinator met with us [the CHWs] in the office and we created the initial set up and user experience for zoom." • "There was steady progress with features such as using the chat box, chatting, and muting participants as the CHWs became more familiar with interface [sic]."
Participants did not know how to download apps on their smartphone	• Call the participant or involve a family member and walk them through the steps to download the app • Provide in-language and empowerment to help participants address frustrations with new technology	• "[We] encourage the participant that a challenge they experience is common (such as problems with the audio), and this encourages the participant to continue to work through the technical situation."
Challenges with scanning consent forms or other survey documents	• Provide real-time virtual or telephone assistance to troubleshoot issues with documents rather than waiting for weekly meetings or follow-up meetings or calls	• "At the time that the CHW/participant needs to send a form, they call the coordinator/CHW to learn the skill. The person receiving the form then sometimes makes quick edits to the form to make it legible or printable."
Participants did not have email addresses to receive study documents	• Involve family members who can provide email addresses or support participants • Choose a communication strategy that the participant might be familiar with, such as a text messaging platform	• "It is often easier to send a link through text message or WhatsApp than email. In general, email proficiency of the participants has not change over time. It is not seen as an easy form of communication in the community."
Participants did not know how to log in to Zoom	• Involve family members in the meetings • Schedule meetings on the basis of the availability of other family members	• "We ask family members when they will be available and try to schedule sessions at those times."
Internet connectivity challenges	• Be flexible with scheduling and offer assistance at multiple times and days of the week, including weekends and evenings to attend sessions • Change to audio only or telephone sessions if participants cannot connect	• "Multiple family members may be using the same smartphone for internet; thus, we offer our sessions on the weekend and evenings and often have to reschedule to make sure participants can attend [sic]."

Moreover, they have addressed several known challenges of group visits, which are both participant- and visit-related. Firstly, the barrier of adequate transportation, particularly acute in large, diffuse urban areas including Atlanta—which has limited public transportation options—can prevent participants from attending group visits. Previous studies on barriers to participation in group medical visits include the lack of transportation or scheduling conflicts [26,27]; nonetheless, by offering this virtual option, participants report that this has facilitated their ability to engage in the intervention. One female CHW commented, "For our participants, Zoom is better, because they don't need to get ready or drive anywhere. We are able to spend more time with them."

Logistical issues of meeting spaces and workflow can lead to inefficiency and be time-intensive for group visit implementation in clinical practice [27,28]. The transition to virtual visits has provided flexibility to the team for scheduling and location, which has likely enhanced participation and engagement and improved the efficiency of the visits. The CHWs report having more flexibility to conduct multiple sessions, if needed, to facilitate participation, as these logistical obstacles are easier to overcome in virtual sessions.

Lastly, all participants share a common background and health conditions. Shared personal health characteristics can greatly enhance participation and patient activation, leading to improved benefits and health outcomes from group visits [26,28]. While

XSL•FO
RenderX

still in progress, the retention rate is >90% toward the fifth month of this 6-month program. At the height of the pandemic, at a time when older adults of underserved racial and ethnic communities with LEP were at the greatest risk of social isolation, the virtual group visits have provided much welcomed social interaction for these at-risk individuals. One female CHW commented, “They [participants] are bored at home; now they are eagerly awaiting our sessions!”

While this pandemic has certainly strained most aspects of well-being for a majority of individuals, there are opportunities to shape our “new normal.” Namely, for communities of color, older adults, and those with LEP, where the risk of the digital divide and social isolation may widen, investing in resources, including CHW-led group virtual visits, may help address gaps in care. The group visit model offers the opportunity to enhance social interaction and patient activation, and preliminary results from our study indicate that the virtual group visit is feasible and acceptable for older adults with LEP. Further, we believe that the involvement of CHWs is critical to this process, to help bridge the digital void that is at risk of widening further during this ongoing pandemic and to enhance the digital literacy of our most vulnerable populations. Beyond the current pandemic, we believe that this model is feasible and can continue to reach our most vulnerable populations in the future.

Acknowledgments

We would like to acknowledge the dedicated work of our CHWs, Sakila Nasrin, Zohra Amin, and Nazneen Akter, who have committed themselves to the success of our study in spite of the challenges faced during the COVID-19 pandemic. We would also like to acknowledge our DREAM partners in New York City for their mentorship and guidance. This study was supported by National Institute on Minority Healthy and Health Disparities (K23 MD015088-01 and 3U54MD000538-17S1). MKA and KMV received support from P30DK111024 awarded by the Georgia Center for Diabetes Translation Research, funded by the National Institute of Diabetes and Digestive and Kidney Diseases. The funders did not play a role in the design, collection, analysis, or drafting of the manuscript. NI is partially supported by the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (grants R01DK110048-01A1, R18DK110740, and P30 DK111022R01DK11048), National Heart, Lung, and Blood Institute (grant 1UG3HL151310), and National Center for Advancing Translational Science (grant UL1TR001445); and by the Centers for Disease Control and Prevention (grant U48DP001904).

Conflicts of Interest

None declared.

References

1. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing Equity in Telemedicine for Chronic Disease Management During the Covid-19 Pandemic. NEJM Catalyst 2020;1(3) [FREE Full text] [doi: 10.1056/cat.20.0131]
2. Roberts ET, Mehrotra A. Assessment of Disparities in Digital Access Among Medicare Beneficiaries and Implications for Telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389. [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
3. Gordon NP, Hornbrook MC. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
4. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a Health Information-Seeking Population and the Current Digital Divide: Cross-Sectional Analysis of the 2015-2016 California Health Interview Survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
5. Chunara R, Zhao Y, Chen J, Lawrence K, Testa P, Nov O, et al. Telemedicine and healthcare disparities: a cohort study in a large healthcare system in New York City during COVID-19. J Am Med Inform Assoc 2021 Jan 15;28(1):33-41 [FREE Full text] [doi: 10.1093/jamia/ocaa217] [Medline: 32866264]
6. Yoon H, Jang Y, Vaughan PW, Garcia M. Older Adults' Internet Use for Health Information: Digital Divide by Race/Ethnicity and Socioeconomic Status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
7. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the Internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
8. Casillas A, Moreno G, Grotts J, Tseng C, Morales LS. A Digital Language Divide? The Relationship between Internet Medication Refills and Medication Adherence among Limited English Proficient (LEP) Patients. J Racial Ethn Health Disparities 2018 Dec;5(6):1373-1380 [FREE Full text] [doi: 10.1007/s40615-018-0487-9] [Medline: 29600351]
9. Czeisler M, Marynak K, Clarke KE, Salah Z, Shakya I, Thierry JM, et al. Delay or Avoidance of Medical Care Because of COVID-19-Related Concerns - United States, June 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1250-1257 [FREE Full text] [doi: 10.15585/mmwr.mm6936a4] [Medline: 32915166]
10. Pollack HA, Kelly C. COVID-19 And Health Disparities: Insights From Key Informant Interviews. Health Affairs Blog. 2020 Oct 27. URL: https://www.healthaffairs.org/do/10.1377/hblog20201023.55778/full/ [accessed 2020-12-15]

XSL•FO
RenderX

11. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health 2020 Aug;8(8):e1003-e1017 [FREE Full text] [doi: 10.1016/S2214-109X(20)30264-3] [Medline: 32553130]
12. CDC COVID-19 Response Team. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 03;69(13):382-386 [FREE Full text] [doi: 10.15585/mmwr.mm6913e2] [Medline: 32240123]
13. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
14. Cunningham SA, Adams SR, Schmittdiel JA, Ali MK. Incidence of diabetes after a partner's diagnosis. Prev Med 2017 Dec;105:52-57 [FREE Full text] [doi: 10.1016/j.ypmed.2017.08.020] [Medline: 28823754]
15. Strom JL, Egede LE. The impact of social support on outcomes in adult patients with type 2 diabetes: a systematic review. Curr Diab Rep 2012 Dec;12(6):769-781 [FREE Full text] [doi: 10.1007/s11892-012-0317-0] [Medline: 22949135]
16. Kobayashi LC, Steptoe A. Social Isolation, Loneliness, and Health Behaviors at Older Ages: Longitudinal Cohort Study. Ann Behav Med 2018 May 31;52(7):582-593 [FREE Full text] [doi: 10.1093/abm/kax033] [Medline: 29860361]
17. Careyva BA, Johnson MB, Goodrich SA, Shaak K, Stello B. Clinician-Reported Barriers to Group Visit Implementation. J Prim Care Community Health 2016 Jul;7(3):188-193 [FREE Full text] [doi: 10.1177/2150131916631924] [Medline: 26883437]
18. Thompson C, Meeuwisse I, Dahlke R, Drummond N. Group medical visits in primary care for patients with diabetes and low socioeconomic status: users' perspectives and lessons for practitioners. Can J Diabetes 2014 Jun;38(3):198-204. [doi: 10.1016/j.jcjd.2014.03.012] [Medline: 24909090]
19. Peretz PJ, Islam N, Matiz LA. Community Health Workers and Covid-19 - Addressing Social Determinants of Health in Times of Crisis and Beyond. N Engl J Med 2020 Nov 05;383(19):e108. [doi: 10.1056/NEJMp2022641] [Medline: 32966715]
20. Jeet G, Thakur JS, Prinja S, Singh M. Community health workers for non-communicable diseases prevention and control in developing countries: Evidence and implications. PLoS One 2017;12(7):e0180640 [FREE Full text] [doi: 10.1371/journal.pone.0180640] [Medline: 28704405]
21. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British South Asian Patients' Perspectives on the Relevance and Acceptability of Mobile Health Text Messaging to Support Medication Adherence for Type 2 Diabetes: Qualitative Study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15789 [FREE Full text] [doi: 10.2196/15789] [Medline: 32310150]
22. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
23. Narasimha S, Madathil KC, Agnisarman S, Rogers H, Welch B, Ashok A, et al. Designing Telemedicine Systems for Geriatric Patients: A Review of the Usability Studies. Telemed J E Health 2017 Jun;23(6):459-472. [doi: 10.1089/tmj.2016.0178] [Medline: 27875667]
24. Ellimoottil C, An L, Moyer M, Sossong S, Hollander JE. Challenges And Opportunities Faced By Large Health Systems Implementing Telehealth. Health Aff (Millwood) 2018 Dec;37(12):1955-1959. [doi: 10.1377/hlthaff.2018.05099] [Medline: 30633667]
25. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth Interventions Delivering Home-based Support Group Videoconferencing: Systematic Review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
26. Wadsworth KH, Archibald TG, Payne AE, Cleary AK, Haney BL, Hoverman AS. Shared medical appointments and patient-centered experience: a mixed-methods systematic review. BMC Fam Pract 2019 Jul 08;20(1):97 [FREE Full text] [doi: 10.1186/s12875-019-0972-1] [Medline: 31286876]
27. Jaber R, Braksmajer A, Trilling JS. Group visits: a qualitative review of current research. J Am Board Fam Med 2006;19(3):276-290 [FREE Full text] [doi: 10.3122/jabfm.19.3.276] [Medline: 16672681]
28. Edelman D, Gierisch JM, McDuffie JR, Oddone E, Williams JW. Shared medical appointments for patients with diabetes mellitus: a systematic review. J Gen Intern Med 2015 Jan;30(1):99-106 [FREE Full text] [doi: 10.1007/s11606-014-2978-7] [Medline: 25107290]

Abbreviations

CHW: community health worker
LEP: limited English proficiency

XSL•FO
RenderX

Edited by R Kukafka; submitted 02.02.21; peer-reviewed by K Baucom, D Petsani; comments to author 26.03.21; revised version received 12.04.21; accepted 14.06.21; published 08.07.21.

Please cite as:
Shah MK, Gibbs AC, Ali MK, Narayan KMV, Islam N
Overcoming the Digital Divide in the Post–COVID-19 "Reset": Enhancing Group Virtual Visits with Community Health Workers
J Med Internet Res 2021;23(7):e27682
URL: https://www.jmir.org/2021/7/e27682
doi:10.2196/27682
PMID:34152995

©Megha K Shah, Ashley Christina Gibbs, Mohammed K Ali, K M Venkat Narayan, Nadia Islam. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Time to Change for Mental Health and Well-being via Virtual Professional Coaching: Longitudinal Observational Study

Alexis M Jeannotte[1], BS, PhD; Derek M Hutchinson[1], BA, MA, PhD; Gabriella R Kellerman[1], BA, MD

BetterUp, Inc, San Francisco, CA, United States

Corresponding Author:
Alexis M Jeannotte, BS, PhD
BetterUp, Inc
1200 Folsom Street
San Francisco, CA, 94103
United States
Phone: 1 7034022166
Email: ajeannotte28@gmail.com

Abstract

Background: Optimal mental health yields many benefits and reduced costs to employees and organizations; however, the workplace introduces challenges to building and maintaining mental health that affect well-being. Although many organizations have introduced programming to aid employee mental health and well-being, the uptake and effectiveness of these efforts vary. One barrier to developing more effective interventions is a lack of understanding about how to improve well-being over time. This study examined not only whether employer-provided coaching is an effective strategy to improve mental health and well-being in employees but also how this intervention changes well-being in stages over time.

Objective: The goal of this study was to determine whether BetterUp, a longitudinal one-on-one virtual coaching intervention, improves components of mental health and psychological well-being, and whether the magnitude of changes vary in stages over time. This is the first research study to evaluate the effectiveness of professional coaching through three repeated assessments, moving beyond a pre-post intervention design. The outcomes of this study will enable coaches and employers to design more targeted interventions by outlining when to expect maximal growth in specific outcomes throughout the coaching engagement.

Methods: Three identical assessments were completed by 391 users of BetterUp: prior to the start of coaching, after approximately 3-4 months of coaching, and again after 6-7 months of coaching. Three scales were used to evaluate psychological and behavioral dimensions that support management of mental health: stress management, resilience, and life satisfaction. Six additional scales were used to assess psychological well-being: emotional regulation, prospection ability, finding purpose and meaning, self-awareness, self-efficacy, and social connection.

Results: Using mixed-effects modeling, varying rates of change were observed in several dimensions of mental health and psychological well-being. Initial rapid improvements in the first half of the intervention, followed by slower growth in the second half of the intervention were found for prospection ability, self-awareness, self-efficacy, social connection, emotional regulation, and a reduction in stress (range of unstandardized β values for each assessment: .10-.19). Life satisfaction improved continuously throughout the full intervention period (β=.13). Finding purpose in meaning at work and building resilience both grew continuously throughout the coaching intervention, but larger gains were experienced in the second half of the intervention (β=.08-.18), requiring the full length of the intervention to realize maximal growth.

Conclusions: The results demonstrate the effectiveness of BetterUp virtual one-on-one coaching to improve psychological well-being, while mitigating threats to mental health such as excessive and prolonged stress, low resilience, and poor satisfaction with life. The improvements across the collection of outcomes were time-dependent, and provide important insights to users and practitioners about how and when to expect maximal improvements in a range of interrelated personal and professional outcomes.

(J Med Internet Res 2021;23(7):e27774) doi:10.2196/27774

KEYWORDS

professional coaching; virtual coaching; mental health; psychological well-being; stress; resilience; life satisfaction; longitudinal; intervention; well-being; satisfaction; coach; observational

XSL•FO
RenderX

Introduction

Background

In recent years, the cost of mental health to organizations worldwide has grown dramatically, with poor mental health becoming one of the leading risks to worker well-being and safety [1]. In the workplace, poor mental health, and its clinical counterpart mental illness, can present significant challenges to individual, team, and organizational functioning, with approximately 19% of working adults in developed countries suffering from a behavioral health disorder [2,3] and almost US $1 trillion dollars in lost productivity to the global economy as a result of depression and anxiety alone [4]. Importantly, many of the symptoms associated with poor mental health do not rise to the level of a clinical diagnosis. Even without reaching this threshold, these symptoms can still be debilitating, or at least disruptive, by impacting the daily lives and professional contributions of those affected [5-9].

Just as poor mental health can cause personal and professional problems, positive psychological well-being at work is associated with a range of beneficial life outcomes. Such benefits include greater happiness, improved health, and increased longevity [10-14]. Improving mental health can also yield benefits to organizations that transcend personal well-being, which include higher levels of job satisfaction, greater organizational commitment, increased work effort, lower employee turnover, lower rates of absenteeism, and fewer workplace accidents [15]. Boehm and Lyobumirsky [16] suggested that happier employees display more citizenship behaviors in the workplace, are better performers, and earn higher salaries, demonstrating tangible benefits for themselves, their colleagues, and their employers.

A multitude of employer-provided interventions have proven to be beneficial in helping employees to both address poor mental health and improve psychological well-being. Such programs include workplace wellness programs, employee assistance programs, mentoring programs, and employee affinity groups [17-22]. The effectiveness of such interventions varies, with usage and efficacy depending on employee uptake, perceptions of stigma around seeking treatment or support, as well as explicit or implicit signals from organizational leadership [22-27]. However, almost nothing is known about how these interventions differentially impact components of well-being over time. This knowledge gap prevents deeper understanding of how these interventions drive change at a level that can ultimately inform the design of even more effective, targeted, and personalized offerings.

In this study, we examined the longitudinal effectiveness of BetterUp, a virtual one-on-one professional coaching platform provided by employers to improve employee mental health and well-being. Professional coaching as an employee benefit is growing in popularity among employers with increasing evidence about its effectiveness to help develop a range of personal and professional mindsets and behaviors to include those that support psychological well-being [28-30]. In particular, BetterUp may overcome the limitations of other employer-provided resources to improve mental health by enabling support that is personalized to the employee, adapts as their needs evolve, and provides access to a coach when and where they need it— whether that be at work or outside of work—avoiding potential stigma or perceptions that seeking support could be viewed negatively by the employer.

We investigated the longitudinal and sequential impact of BetterUp on a set of positive mental health outcomes and related dimensions that help mitigate poor mental health. As shown in Table 1, the measured outcomes were aligned to six dimensions of positive psychological well-being [31] and three dimensions that help to bolster positive mental health, while mitigating negative mental health: stress management [32], resilience [33], and life satisfaction [34]. Critically, maintenance or improvement of each of these dimensions can directly impact personal and professional success [35].

XSL•FO
RenderX

Table 1. Alignment of study outcomes and associated measurement scales to dimensions from prior models of positive mental health and psychological well-being.

Positive mental health and psychological well-being dimensions		BetterUp dimensions			
Dimension	Definition	Dimension	Reliability (Cronbach α)	Items (n)	Sample item
Autonomy [31]	Self-determining and independent; regulates behavior from within	Emotional regulation	.84	3	I have effective strategies for maintaining control of my emotions
Environmental mastery [31]	Sense of competence in managing the context around them	Self-efficacy	.84	3	I believe I can achieve the things that I really want in life
Self-acceptance [31]	Positive attitude toward the self, including good and bad aspects	Self-awareness	.70	4	I have a good sense of the things in life that bring me joy
Purpose in life [31]	Holds beliefs that give life purpose; feels meaning in present and past life	Purpose and meaning	.92	3	The work I do makes an impact
Personal growth [31]	Motivation for continued growth and development	Prospection	.87	3	I think about how to make the most out of my future
Positive relations with others [31]	Satisfying and trusting relationships, understands give and take	Social connection	.80	3	I regularly interact with people who give me support and encouragement
Stress [32]	Experience of tension as a result of personal or work circumstances	Stress	.89	3	I experience a great deal of tension in my daily life
Resilience [33]	Cope, recover, and grow from challenging circumstances	Resilience	.88	3	I recover quickly after stressful experiences
Life satisfaction [34]	Feeling of fulfillment in life	Life satisfaction	.84	2	To what extent are the things you do in your life worthwhile?

Objectives

The goal of this study was to determine whether BetterUp, a longitudinal one-on-one virtual coaching intervention, improved components of mental health and psychological well-being, and whether the magnitude of changes varied over time. Toward this end, we examined the set of outcomes at three different time points to map out the developmental timeline of the professional coaching intervention provided by BetterUp. We surveyed identical outcomes before coaching began, at baseline, and then again at two subsequent time points after coaching began to understand if coaching impacts specific mental health and well-being outcomes, when significant changes occur, and whether improvements are maintained over time.

To our knowledge, this is the first time the effectiveness of professional coaching has been studied with more than two time points (ie, a single baseline and one postcoaching assessment). In particular, our novel use of multiple, repeated assessments to understand the length of time it takes to maximize the benefits of professional coaching for employees enables gaining unique insight into the timeline to slow or prevent a decline in mental health (eg, increasing stress) or bolster positive psychological well-being in individuals. The outcomes of this study will enable coaches and employers to design more targeted interventions by determining when to expect maximal growth in specific outcomes throughout the coaching engagement.

Methods

Study Design and Participants

This study used a longitudinal, observational within-subjects design to evaluate individual changes in the personal and professional mental health and well-being outcomes of BetterUp users. Although users are continuously engaged on the BetterUp platform, we a priori set the data collection window for approximately 1 year to allow sufficient time for users to complete the program. For this study, program completion was defined as completing at least eight coaching sessions, an initial baseline assessment, additional assessments every 3-4 months across two time points, and remaining active on the platform for approximately 6 months. Participation on the platform was voluntary and individuals had the option to stop using the services at any time.

The use of human subject data was reviewed and determined to be exempt by the Ethical and Independent Review Services Institutional Review Board.

Intervention

BetterUp provides one-on-one virtual coaching to support an individual's personal and professional development. The platform provides coachees access to over 2000 coaches with the coach/coachee fit optimized by a set of algorithms that make recommendations based on the coachee's personal and professional characteristics, motivations, and interests (Figure 1). Each coaching session is conducted primarily via video with a professionally certified coach and lasts approximately 30-45 minutes (Figure 1). Although the majority of the

XSL•FO
RenderX

communications occur via video, the platform also allows for text messaging and phone communications.

Figure 1. BetterUp Platform. The BetterUp virtual coaching platform offers users an algorithm-driven coach-selection process (left), coaching sessions primarily via video (middle), and resources for further personal development resulting from hybrid coach and algorithm recommendations (right).



Users of BetterUp complete an assessment of their mental health, well-being, and professional attributes and goals prior to starting the program. They complete these assessments again at approximately 3 to 4-month intervals to evaluate progress. Importantly, no demographic information is collected at any time. The user community is global, coming from countries with varying information privacy and protections requirements. BetterUp chooses not to collect demographic information to maintain privacy and protect the data of its users regardless of their geographical location.

Once the longitudinal coaching engagement is initiated, coaches select and push resources to coachees to further their development outside the coaching session. An algorithm recommends resources to the coach based on the topic of a coaching session, which include guided or self-guided exercises, readings, and audio or video content (Figure 1). Additionally, the platform provides the coachee a set of features to personalize their experience, to include setting and tracking goals, receiving nudges to support behavior change, and joining group-based sessions.

The type of coaching offered by BetterUp adheres to a traditional model of coaching; that is, a "one-to-one learning and development intervention that uses a collaborative, reflective, goal-focused relationship to achieve professional outcomes that are valued by the coachee" [28]. The defining practices of the coaches are the activities of listening, asking questions, empowering, and encouraging goal-setting, in contrast to providing specific recommendations or advice [36-38]. During a typical session, the coachee will take the lead in defining discussion topics, allowing them control over their own development and growth at a comfortable pace [39]. Furthermore, coaching is a practice distinct from clinical care such as therapy, in that it is future-focused, rather than an exploration and understanding of past experiences [40,41]. Coaching provides a resource to support the development of beneficial mindsets and sustained behavior change, which are necessary to mitigate negative mental health and support positive well-being to potentially protect against worsening symptoms and the need for clinical care.

Measure Development and Validation

The primary outcomes used to evaluate individual growth align with a theory-based conception of positive mental health, as distinct from mental illness, and well-being operationalized in a set of self-report survey measures developed by Carol Ryff and colleagues [31,42]. This model defines six dimensions of well-being: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Further development of this model of mental health identified three other important dimensions of psychological well-being that both bolster positive mental health and buffer against negative or declining mental health: stress management [32], resilience [33], and life satisfaction [34].

The existing scales to evaluate these nine dimensions had to be adapted to be used as part of BetterUp for three primary reasons: (1) the original short versions of the existing scales had poor

XSL•FO
RenderX

psychometric properties in some cases, and the 7-item version was too lengthy for commercial purposes [43,44]; (2) the scales and items needed to be framed to assess how work and the workplace shaped employee perceptions of these dimensions; and (3) BetterUp requires commercial usage of the scales and does not qualify exclusively as use for research purposes. As a result, the set of scales that measure these nine dimensions were newly developed for BetterUp and validated in a large multiphase study performed by the coaching provider [45]. The development and validation of these scales are briefly described here, with supporting documentation provided in Multimedia Appendix 1.

Initial items were developed using related measures from the literature and interviews with stakeholders. The full set of items was validated using a large sample of working professionals (N=1030) recruited from the online Amazon Mechanical Turk platform. All initial items were evaluated using various classical test theory item statistics, which included means, SD, skewness, kurtosis, interitem correlations, item-total correlations, and Cronbach α (>.70). Each scale was trimmed down from 3-10 items to 1-4 items, and all scales used in this study were composed of 3-4 items. Final scale reliabilities are reported below for each of the selected scales used in this study.

Convergent validity was determined using measures previously validated in the leadership and well-being literature. For the selected measures in this study, these included the Authentic Leadership Questionnaire [46], Psychological Capital [47], PERMA (positive emotion, engagement, relationships, meaning, accomplishments) [48], self-regulation [49], life satisfaction [50], and meaning and purpose [51]. In unpublished work utilizing a cohort of approximately 1000 US professionals, we also examined the convergent validity of a subset of the measures in this study with a brief screening tool for depression and anxiety, the Patient Health Questionnaire 4 (PHQ-4) [52,53]. Specifically, we found significant negative correlations between PHQ-4 and life satisfaction (r_t –0.390, P<.001), prospection (r_t –0.201, P<.001), purpose and meaning (r_t –0.192, P<.001), resilience (r_t –0.326, P<.001), and social connection (r_t –0.326, P<.001). Additionally, other outcomes not measured in this study also had significant negative relationships with PHQ-4, including happiness and optimism (r_t –0.422 and –0.295, respectively; P<.001).

Multivariate analyses of variance (ANOVAs) were performed to detect differences across managerial status (ie, people leader or individual contributor), gender, and ethnicity on each scale. All variables demonstrated linearity and acceptable univariate normality. Finally, a test-retest assessment was performed at 1 month to assess the reliability of each outcome over time. The test-retest correlations were moderate to high at 1 month (r_s=0.64-0.85), reflecting acceptable and minimal change over time in a noncoached sample.

Reliabilities and example items for each of the nine scales are listed in Table 1. Except where noted, each outcome was measured using a set of three items that were rated by participants for their level of agreement using a 1-5 Likert-type scale.

Statistical Analyses

The within-subject means of item-level responses were determined for each of the outcomes. The means and SDs within the sampled population are reported for each outcome. The Pearson product-moment correlation coefficient was determined for each outcome pairing along with the 95% CI around the coefficient.

To evaluate the change in the outcome measures, our repeated-measures data were evaluated through a series of linear mixed-effect multilevel models using the lme4 package in R [54]. Multilevel modeling allows for the evaluation of both fixed and random effects when there exists substantial nonindependence in observations. This technique represents a more robust statistical approach than repeated-measures ANOVA [55]. All analyses included random intercepts for subject ID to control for the correlation between repeated measurement of the same subjects. The sole predictor in each model was the assessment number that represents the time point during the longitudinal coaching journey that each assessment was administered. Specifically, the baseline (before beginning coaching) assessment was denoted as T1, the second assessment was denoted as T2, and the third and final assessment was denoted as T3.

We first tested if there were significant differences in growth across study measures based on the time point (ie, T1, T2, T3). To evaluate this effect, we constructed a multilevel model assessing the cross-level interaction between a factor composed of our measures and the time point. We report an ANOVA summary statistic based on the fixed effects of the model using the Type III Kenward-Rogers method of approximation produced by the lmerTest package in R [56]. For this specific analysis, our stress measure was reverse-coded to eliminate the chance that our results would be overly skewed by this sole variable decreasing over time, while all others increased over time.

To visualize the change across time, we calculated the group mean–centered effect for each outcome. Scores at each time point were calculated by subtracting an individual's mean across all three time points from their score at each time point. Calculation of the group mean–centered effect allows for a clearer evaluation of the pattern of growth for each coaching outcome by removing variance between individuals.

To examine the magnitude of change of each outcome over time, we calculated the effect size (Cohen d) of each variable for each time period (T1 to T2 and T2 to T3). We then calculated the growth ratio by dividing the effect size from T1 to T2 by the effect size from T2 to T3 to derive an understanding of relative growth across the full coaching engagement. We then applied statistical testing to assess the differences in magnitude of growth between the two time periods. We performed a post hoc test for each outcome measure using the Ken-Roger method of approximation and the Tukey adjustment.

All statistics are reported with a threshold of P<.05 considered for statistical significance; where noted, corrections for multiple comparisons were used. The source code to replicate the statistical analyses is included in Multimedia Appendix 2. The

XSL•FO
RenderX

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

Results

Cohort Characteristics

We included BetterUp users who completed the program between July 15, 2019 and June 24, 2020. The average completion time of the program was 184 days, or approximately 6 months. The exact length of the program varied, as the coachees had flexibility in how often they scheduled coaching sessions and their personal development goals. We set broad inclusion criteria, as described in the Methods section, and had no additional individual exclusions. The final cohort consisted of 391 participants, representing 33% of those who started the program and met the inclusion criteria. A series of independent-sample *t* tests were performed on the initial baseline (ie, T1) assessment across all study variables. No significant differences were found between those who opted into completing follow-up assessments and those who did not. Additionally, of those who did not complete the program, 49% dropped out between T1 and T2, while 51% dropped out between T2 and T3. Data from all 391 participants were used in mixed linear effects modeling.

The BetterUp platform does not collect standard demographic data; therefore, we do not report any results that consider age, gender, or other personal demographic variables. Participants were native English speakers, primarily working and living in the United States. These individuals represented a range of organizations and industries, including professional services, financial services and banking, technology, hospitality, retail, and manufacturing. Among the 391 participants, 75% (n=293) self-identified as managers, while 25% (n=98) identified as individual contributors. This cohort reflects a varied cross-section of workers in the United States today but was not selected to match the exact demographics of this working population.

Multilevel Modeling of Outcomes

A summary of the means (SD) and intercorrelations along with their 95% CIs for all study measures are reported in Table 2; intercorrelations were all significant at $P<.001$. Each measure was aggregated within each individual across all three time points.

The average number of days from T1 to T2 was 97 (SD 42) and the average number of days from T2 to T3 was 87 (SD 49). The variability between assessment points reflected individual variability in schedules and availability to complete coaching sessions, a feature of studying coaching in an applied setting with busy, working professionals.

We first tested if there were significant differences in growth across study measures based on the time point the assessment was completed. As reported in Table 3, we found a significant interaction between the measure factor and the time point. This finding suggests that growth across the assessment time points depends on the specific measure.

The mixed-effect models across all coaching outcome measures are summarized with unstandardized β weights in Table 4. For all models, the intraclass correlation coefficient was greater than 0.50, which suggested substantial variance between individuals across all outcomes. For all variables, there was a significant effect of assessment number on each outcome variable. Our results indicate significant growth during the coaching engagement for the entire set of outcomes (all $P<.001$). The marginal R^2 values represent the proportion of variance explained by the assessment time point [57].

The group mean–centered effects for each variable across time are shown in Figure 2. By removing the variance between individuals, the pattern of growth for each coaching outcome is more clearly observed.

XSL•FO
RenderX

Table 2. Means and correlations of outcome measures with confidence intervals.[a]

Variable	Emotional regulation	Life satisfaction	Prospection	Purpose and meaning	Resilience	Self-awareness	Self-efficacy	Stress
Emotional regulation: mean 3.72 (SD 0.64)								
r	1	0.31	0.19	0.18	0.46	0.28	0.26	–0.30
95% CI	__[b]	0.21 to 0.39	0.09 to 0.28	0.08 to 0.27	0.38 to 0.54	0.19 to 0.37	0.17 to 0.35	–0.39 to –0.21
Life satisfaction: mean 4.08 (SD 0.58)								
r	0.31	1	0.44	0.50	0.39	0.56	0.46	–0.51
95% CI	0.21 to 0.39	—	0.36 to 0.52	0.42 to 0.57	0.30 to 0.47	0.48 to 0.62	0.38 to 0.54	–0.58 to –0.44
Prospection: mean 3.91 (SD 0.59)								
r	0.19	0.44	1	0.32	0.29	0.64	0.44	–0.27
95% CI	0.09 to 0.28	0.36 to 0.52	—	0.23 to 0.40	0.20 to 0.38	0.58 to 0.70	0.36 to 0.52	–0.36 to –0.18
Purpose and meaning: mean 4.16 (0.60)								
r	0.18	0.50	0.32	1	0.26	0.37	0.37	–0.26
95% CI	0.08 to 0.27	0.42 to 0.57	0.23 to 0.40	—	0.17 to 0.35	0.28 to 0.46	0.28 to 0.45	–0.35 to –0.17
Resilience: mean 3.75 (SD 0.63)								
r	0.46	0.39	0.29	0.26	1	0.36	0.43	–0.43
95% CI	0.38 to 0.54	0.30 to 0.47	0.20 to 0.38	0.17 to 0.35	—	0.27 to 0.44	0.35 to 0.51	–0.51 to –0.35
Self-awareness: mean 3.80 (0.56)								
r	0.28	0.56	0.64	0.37	0.36	1	0.41	–0.38
95% CI	0.19 to 0.37	0.48 to 0.62	0.58 to 0.70	0.28 to 0.46	0.27 to 0.44	—	0.32 to 0.49	–0.46 to –0.29
Self-efficacy: mean 4.25 (SD 0.52)								
r	0.26	0.46	0.44	0.37	0.43	0.41	1	–0.28
95% CI	0.17 to 0.35	0.38 to 0.54	0.36 to 0.52	0.28 to 0.45	0.35 to 0.51	0.32 to 0.49	—	–0.37 to –0.18
Stress: mean 2.94 (SD 0.66)								
r	–0.30	–0.51	–0.27	–0.26	–0.43	–0.38	–0.28	1
95% CI	–0.39 to –0.21	–0.58 to –0.44	–0.36 to –0.18	–0.35 to –0.17	–0.51 to –0.35	–0.46 to –0.29	–0.37 to –0.18	—

[a]The CI is a plausible range of population correlations that could have caused the sample correlation [58]; all correlations are significant at $P<.001$.
[b]Not applicable.

Table 3. Outcome aggregated effects by assessment time point.[a]

Predictor	Sum of squares	Mean square	*df* (numerator)	*df* (denominator)	*F*	*P* value
Assessment time point	150.79	150.79	1	10149	459.40	<.001
Coaching outcome	482.45	60.31	8	10149	183.74	<.001
Assessment time point× coaching outcome	9.42	1.18	8	10149	3.59	<.001

[a]In this analysis, stress was reverse-coded to allow for a direct comparison by eliminating the chance our results would be overly skewed by stress being the one variable that decreases over time.

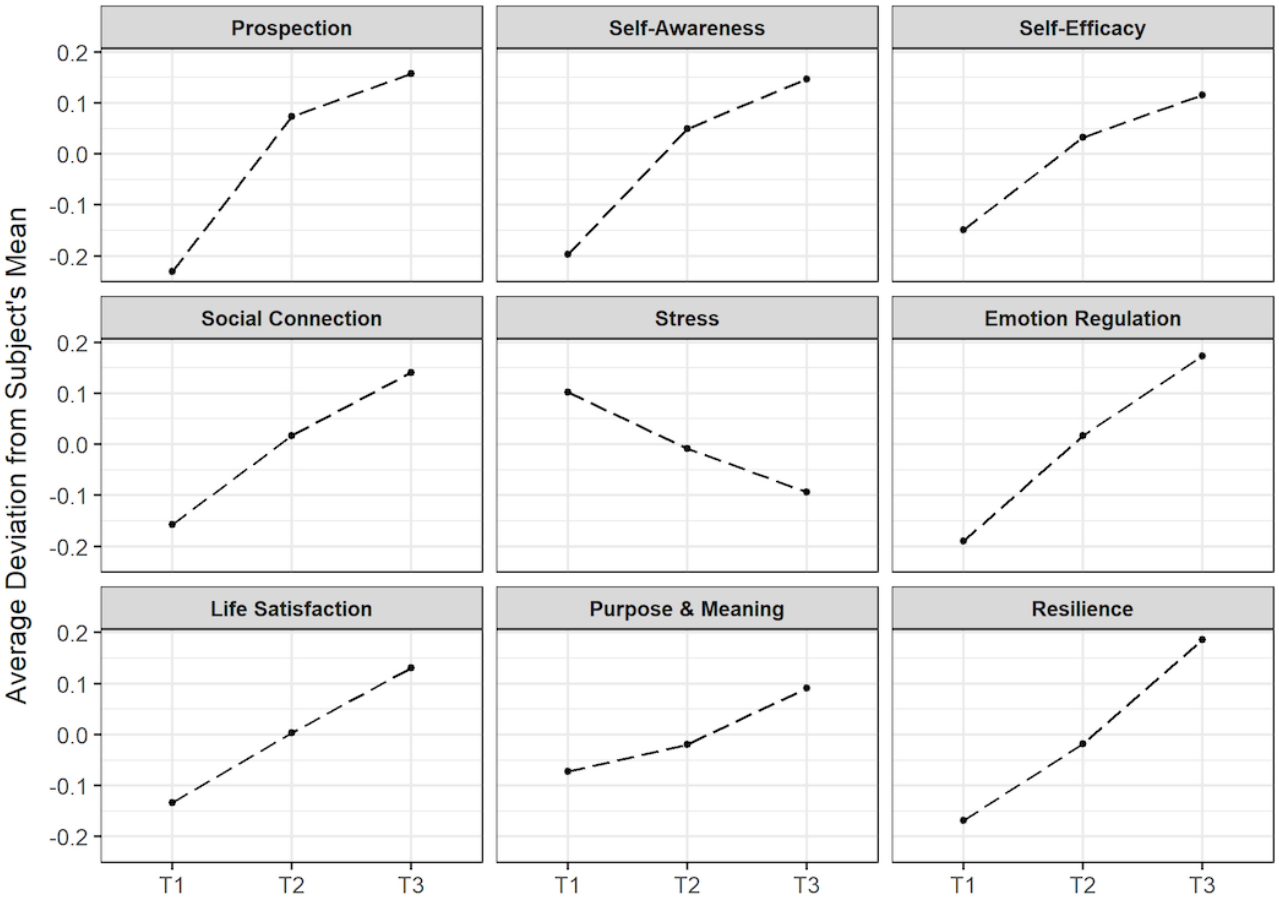
XSL•FO
RenderX

Table 4. Multilevel modeling regression by outcome.

Predictor	Emotional regulation	Life satisfaction	Prospection	Purpose and meaning	Resilience	Self-awareness	Self-efficacy	Social connection	Stress
Intercept (95% CI)	3.53 (3.46 to 3.61)	3.95 (3.89 to 4.02)	3.72 (3.65 to 3.78)	4.08 (4.01 to 4.15)	3.57 (3.50 to 3.64)	3.63 (3.57 to 3.70)	4.11 (4.06 to 4.17)	3.95 (3.90 to 4.00)	3.04 (2.96 to 3.11)
Assessment time point, β (95% CI)	.18 (.15 to .21)	.13 (.10 to .16)	.19 (.16 to .23)	.08 (.05 to .11)	.18 (.15 to .21)	.17 (.14 to .20)	.13 (.11 to .16)	.15 (.12 to .18)	–.10 (–.14 to .06)
Random effects									
σ2	0.22	0.16	0.20	0.18	0.18	0.14	0.13	0.14	0.30
τ00 (ID)	0.34	0.28	0.28	0.30	0.34	0.27	0.23	0.16	0.34
ICC[a]	0.60	0.64	0.59	0.62	0.65	0.66	0.64	0.54	0.53
Marginal R^2	0.038	0.026	0.049	0.009	0.039	0.046	0.032	0.046	0.010

[a]ICC: intraclass correlation coefficient.

Figure 2. Outcome-level growth trajectories over time. Scores for each outcome measure are shown as group mean–centered effects at baseline (T1), during the intervention (T2), and at the end of the intervention (T3).



Longitudinal Growth Trajectories

The magnitude of growth over time and the growth ratios for each outcome are shown in Table 5. The growth ratios are ordered from the largest relative growth during the first time period (T1 to T2) relative to the second time period (T2 to T3) through to the largest relative growth during the second time period relative to that of the first. A growth ratio greater than 1.0 indicates that larger relative growth occurred between T1 and T2 than between T2 and T3 (ie, rapid initial growth), whereas a growth ratio less than 1.0 indicates that larger relative growth occurred between T2 and T3 (ie, delayed growth). A growth ratio of 1.0 indicates linear and sustained growth across both periods of time.

XSL•FO
RenderX

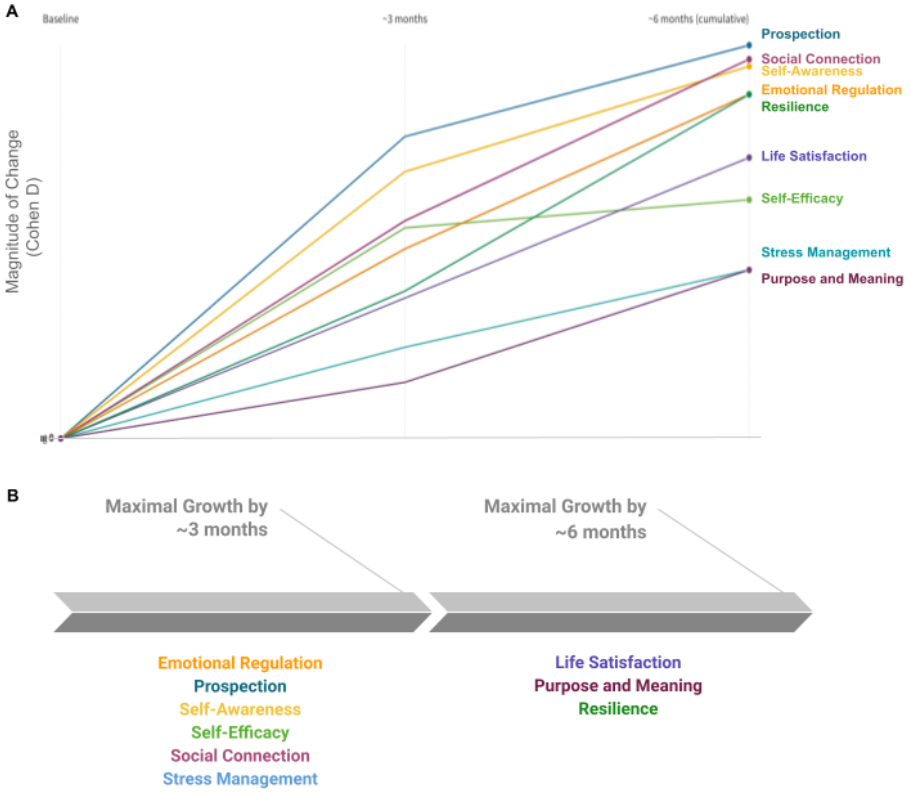
Table 5. Longitudinal growth ratios ordered by effect size.

Variable	Cohen *d*		Growth ratio
	T1–T2	T2–T3	
Prospection	0.43	0.13	3.41
Self-awareness	0.38	0.15	2.47
Self-efficacy	0.30	0.14	2.09
Social connection	0.30	0.24	1.27
Stress	0.13	0.11	1.24
Emotional regulation	0.27	0.22	1.22
Life satisfaction	0.20	0.20	1.01
Resilience	0.21	0.28	0.73
Purpose and meaning	0.08	0.16	0.48

To compare the relative growth of each dimension within each time period and over the cumulative coaching experience, Figure 3 visualizes the changes in two different ways. In this figure, stress is reverse-coded and denoted as "stress management" to enable a better comparison with the other outcomes. Additionally, each dimension is shown in a different color to clearly differentiate change patterns over time.

Figure 3A shows the cumulative effect sizes (ie, magnitude of growth) in each measured outcome between assessment time points, denoted as change at approximately 3 months and again at approximately 6 months, respectively. From this figure, it is evident that BetterUp users experienced the largest growth in prospection throughout the coaching intervention. Moreover, this growth was the largest during the first time period from T1 to T2 (measured at approximately 3-4 months after coaching began) and continued to grow, but by a smaller amount, during the second time period (approximately 6 months after coaching began). Conversely, the smallest and slowest growth, albeit significant, is seen for purpose and meaning. Finally, Figure 3 shows the relatively consistent growth in stress management, life satisfaction, emotional regulation, and social connection, although the magnitude of growth in each of these dimensions varies.

Figure 3. Growth magnitude throughout the intervention. (A) The effect size for each outcome from T1 to T2 and T2 to T3 is shown as the change from zero (T1) prior to the coaching intervention to approximately 3 months after beginning the intervention, and then the cumulative effect size change after approximately 6 months in the intervention. (B) Each outcome measure is separated into the time period, T1 to T2 at ~3 months or T2 to T3 at ~6 months, during which the outcome experienced the largest change.



XSL•FO
RenderX

We also sought to identify the point at which each dimension experiences its maximal change due to the coaching intervention. This result allows a practitioner to understand the minimal amount of time to invest in coaching to experience the largest growth, on average, in a particular dimension of positive mental health and psychological well-being. As shown in Figure 3B, six of the nine measured dimensions showed the largest growth in approximately 3 months of BetterUp coaching, from T1 to T2: emotional regulation, prospection, self-awareness, self-efficacy, social connection, and stress management. Each of these six dimensions continues to increase in the subsequent 3 months; however, the magnitude of growth was lower in the second time period. The remaining three dimensions, resilience, life satisfaction, and feelings of purpose and meaning, showed continued or greater growth between about 3 and 6 months of BetterUp coaching (T2 to T3). This trend suggests that these dimensions benefit from greater investment of time and work to fully develop.

Post-hoc Analyses

The results of post-hoc analyses are reported in Table 6. Briefly, we observed a significant decrease in stress between T1 and T2 and no significant change between T2 and T3. Conversely, the purpose and meaning dimension only significantly increased between T2 and T3, but not from T1 to T2. The remaining outcomes showed significant growth in both time periods; however, as shown in Figure 1 and Table 5, the magnitude of these changes varied by dimension.

Table 6. Results of post-hoc analyses.

Outcome	Mean difference (SE)[a]	*t* (*df*=782)	*P* value
Emotional regulation			
T1–T2	–0.205 (0.034)	–6.056	<.001
T2–T3	–0.158 (0.034)	–4.674	<.001
Life satisfaction			
T1–T2	–0.134 (0.029)	–4.700	<.001
T2–T3	–0.129 (0.029)	–4.521	<.001
Prospection			
T1–T2	–0.305 (0.032)	–9.685	<.001
T2–T3	–0.084 (0.032)	–2.671	.02
Purpose and meaning			
T1–T2	–0.052 (0.030)	–1.705	.20
T2–T3	–0.111 (0.030)	–3.634	.001
Resilience			
T1–T2	–0.151 (0.030)	–4.985	<.001
T2–T3	–0.203 (0.030)	–6.693	<.001
Self-awareness			
T1–T2	–0.245 (0.026)	–9.340	<.001
T2–T3	–0.096 (0.026)	–3.673	<.001
Self-efficacy			
T1–T2	–0.179 (0.025)	–7.098	<.001
T2–T3	–0.083 (0.025)	–3.263	.003
Social connection			
T1–T2	–0.174 (0.027)	–6.511	<.001
T2–T3	–0.124 (0.027)	–4.637	<.001
Stress			
T1–T2	0.110 (0.039)	2.812	.01
T2–T3	0.087 (0.039)	2.223	.07

[a]Paired contrasts of estimated marginal means controlling for between-individual variance using the latter group as the comparison group; therefore, a negative estimate indicates that the outcome at T1 is lower than the outcome at T2, or that the outcome at T2 is lower than that at T3.

Overall, our results demonstrate that there are different growth rates across a range of mental health and psychological well-being dimensions during a longitudinal coaching engagement. These results provide evidence of substantial

XSL•FO
RenderX

growth in personal and professional outcomes throughout multiple months of one-on-one virtual coaching, and that the growth in these variables is dynamic as an individual develops various mindsets and behaviors through the BetterUp intervention.

Discussion

Main Findings

In this study, we demonstrated that BetterUp's one-on-one virtual coaching improved positive mental health and psychological well-being in as short as 3 months, and drove continued benefits when users were engaged in professional coaching over longer periods of time. This is the first study to examine and establish the time-dependent and multiphase changes in specific dimensions of positive mental health and psychological well-being through an employer-sponsored coaching program.

This study used repeated assessments at baseline, part way through the coaching intervention, and after completing the intervention to identify a rich time course of personalized growth. These detailed trajectories can better inform coaches, coachees, and employers about the timing and magnitude of the mental health and well-being changes they can expect from a coaching intervention. In our examination of a range of factors that comprise positive mental health and psychological well-being (self-awareness, self-efficacy, emotional regulation, prospection, purpose and meaning, and social connection) and those that are important for managing mental health (stress, resilience, and life satisfaction), we demonstrated a complex interplay between these variables with improvements in each occurring at different rates during a longitudinal coaching engagement.

Coaching to Mitigate Poor Mental Health and Bolster Positive Mental Health

Most notably, we found immediate significant reductions in stress within approximately 3-4 months of coaching (T1 to T2) that numerically continued to decrease over time, but without further significant improvements. Simultaneously, there was significant and linear growth in life satisfaction, and a slightly delayed, yet significant, increase in resilience throughout the full coaching program. Together, development of these three dimensions demonstrates the potential role of coaching to reduce the incidence of poor mental health, while supporting the development of resources to enable longitudinal mental health and well-being [32-34,59]. The different growth rates also suggest that individuals should first focus on reducing stress before expecting ongoing improvements in life satisfaction or resilience. Additional work should explore whether the immediate and significant reductions in stress are necessary to detect continued growth in the other measured outcomes, thereby delineating whether effective management of stress with the help of a coach acts as a gateway to further growth in other dimensions of mental health.

Given the personal, organizational, and societal costs of poor mental health, our results demonstrate the broad and far-reaching positive impacts of personalized professional leadership coaching. The results demonstrated that the investment of time and resources in an activity such as personalized professional coaching buffers individuals from the daily personal and work stressors that can erode mental health over time; allows them to build the skills to adapt, recover, and grow from challenging situations; as well as improves their outlook on their life situation. Although the coaching engagement was not directed specifically at improving mental health, the behaviors and mindsets developed are foundational for healthy mental functioning in all aspects of life.

Longitudinal Growth in Positive Psychological Well-being Through Coaching

We found the largest growth from T1 (baseline) to T2, relative to that occurring from T2 to T3, in prospection abilities, self-awareness, and self-efficacy. Collectively, these three dimensions capture an individual's ability to understand the factors that will influence their own development through the coaching experience. Prospection involves an individual using their past experiences and current goals to plan for the future, to include identifying potential obstacles [60-62]. The visioning and goal-setting that define prospection would explain the early and rapid growth in this dimension with ~3-4 months of coaching, as these activities are core to the coaching relationship and start early in the coaching engagement for building the motivation to work on the mindsets and behaviors that will aid goal attainment. Self-awareness measures the conscious attention to focus on the self, whereas self-efficacy captures the belief that an individual can achieve their goals. These dimensions complete a motivational framework for goal-setting and goal achievement. Together, the large early growth in these three dimensions suggests that coaches are guiding individuals in an introspective assessment of their needs and goals (self-awareness), developing goals and the plan for achieving them while identifying potential barriers (prospection), and building the beliefs that this vision is attainable (self-efficacy). Thus, significant early improvement in these three dimensions of psychological well-being are foundational elements of the intentional mindset and behavioral change needed to support positive mental health.

We found significant growth in the two remaining dimensions of psychological well-being, emotional regulation and social connection, throughout the intervention, but with slightly larger growth in the first 3-4 months (T1 to T2) relative to the change observed from approximately 3-6 months (T2 to T3). Emotional regulation, which aligns with the autonomy dimension of the Ryff model of psychological well-being [31], defines how well we internally regulate our emotions, thoughts, and behaviors. Importantly, such regulation promotes resilience [63], which is evident in our study given the greatest amount of growth experienced with a lengthier intervention of approximately 3-6 months of coaching. The link between emotional regulation and resilience should be further explored to understand whether improved emotional regulation is a necessary precursor for increased resilience. Additionally, social connection, which involves building and maintaining close and trusting relationships, is a behavior that experiences continual growth with the help of a coach as the coachee evolves and shifts how they may engage in interpersonal interactions. Importantly,

XSL•FO
RenderX

social connection itself is strongly correlated with well-being and with dimensions that mitigate negative mental health such as stress and resilience [64,65], indicating that continued growth in this dimension through coaching supports positive mental health.

Finally, we found that the psychologically demanding processes of finding purpose and meaning, as well as building resilience took the longest time to fully mature. Specifically, the increase in these dimensions from T2 to T3 was approximately 50% and 30% greater than that from T1 to T2, respectively (Table 5 and Figure 3A).

In particular, finding purpose in life is a complex dimension that includes developing and holding beliefs about meaning of self, behaviors, pursuits, and others both in the past and present. We found that BetterUp users experienced significant growth in this dimension, but it took the entire length of the intervention, approximately 6 months, to experience maximal change. Notably, this study ended after the third assessment; however, future work should examine whether the development of this mindset, and others, continues to increase or plateaus.

Across the range of outcomes, we found specific groupings of dimensions that improved at similar rates. Such groupings suggest an underlying link in the psychological, cognitive, or affective processes that support dimensions that change at similar rates. Although we were not able to test this causal hypothesis directly, such a question could be pursued in future experimental studies with a prospective design that includes a matched control group.

Professional Coaching to Boost Mental Health and Well-being

The patterns of growth identified in this study indicate that the model of one-on-one virtual coaching enables improvement in certain mindsets and behaviors before others, which act as building blocks to enable later growth in other constructs. This trend is most clearly seen in two different dimensions: (1) in the immediate and significant reduction in stress after approximately 90 days of engaging with a professional coach and the maintenance of this lower stress over the subsequent 3 months; and (2) the delayed growth in resilience, as well as purpose and meaning, which both significantly increased after approximately 6 months of coaching, highlighting the benefit of continued engagement with a coach to achieve growth in an area with demonstrated ties to mental health and well-being [66].

It is also notable that the measures that were used in this study were not clinical assessments; thus, they reflect an evaluation of psychological, cognitive, and affective states that may provide early indicators of growing maladaptive behaviors or, conversely, improvements that support mental flourishing that should be of interest to employers. However, in unpublished work (see Methods for more details), we identified a significant negative correlation between the PHQ-4, a brief screening tool for depression and anxiety, with resilience, life satisfaction, prospection, social connection, and meaning and purpose. This suggests that by improving these outcomes, there is an increased likelihood of a concomitant reduction in symptoms of depression and anxiety that may prevent a worsening of mental health. By capturing the effectiveness of longitudinal coaching to improve all of these outcomes, we suggest that an intervention such as one-on-one coaching can actually stop, or even reverse, the worsening of mental health. In particular, the significant reduction in stress and the continued increase in resilience while working with a professional coach focused on leadership development highlight that virtual one-on-one coaching can be used to mitigate a decline in mental health, while simultaneously building the skills to maintain or even improve mental health during challenging times, even when this is not the intended goal.

Limitations and Future Work

Despite our finding of growth in various dimensions of mental health and well-being for BetterUp users, we recognize a few limitations of this study. Because of the retrospective and observational study design, we are unable to draw a causal link between the role of coaching on the different outcome measures. Subsequent research could include a noncoached control group that is assessed over the same period of time without any intervention and a noncoached control group that is provided self-guided materials intended to develop the same outcomes.

Our study design also required us to focus on individual outcomes separately, rather than looking at the temporal interplay between them. Such an analysis would have required a group that did not receive coaching, along with measures taken at higher frequency to enable a finer-grained time-series analysis with the appropriate level of statistical power. This type of design would also demand fewer outcome measures to maintain robust power while examining all interactions. Therefore, we consider this study to provide the first evidence of the specific outcomes that are more or less amenable to change through coaching and at what time point these changes would occur in the process. In particular, prospection, as well as purpose and meaning, showed the extremes of rapid and delayed growth, respectively. Given the literature demonstrating the impact of poor mental health on organizations, the shifts in stress and resilience through employer-provided coaching may have widespread impacts on workforce engagement and productivity. These outcomes may provide the best targets for follow-up studies.

The mixed linear effects modeling approach employed in this study allowed us to examine the variance attributed to the fixed effects relative to random effects. Given our ability to partial out the fixed effects, it should be noted that the effect of the assessment time points was relatively small, which suggests that a myriad of other factors explain the change in each of the dimensions of well-being and mental health. These factors likely stem from the nature of a typical one-on-one coaching journey, which is highly individualized to the needs and goals of the coachee. Additional research should explore the impact of individual psychological, cognitive, affective, or social factors; different rates of learning; or differential success in practicing coached dimensions in everyday life to further explain the mechanism by which one-on-one coaching is effective.

The decision to examine a cohort that was actively engaged in professional coaching through a virtual platform introduced two

XSL•FO
RenderX

additional constraints: the use of self-report measures and a highly motivated population that is more likely to experience greater levels of growth. With respect to the latter point, we actually believe that this is a strength of the applied setting, as this enabled observing real-life growth, while still capturing the varying uptake and impact across participants. With respect to the former point, future studies could seek out populations that would allow for physical access, thus enabling the direct observation of behavioral changes. Additionally, by taking the study out of a very applied context, researchers could examine the effect of varying levels of motivation and the interactions between different outcome measures. Such study elements would allow for gaining a deeper understanding of how individuals experience coaching and when coaching is most efficacious.

Conclusions

In summary, we found distinct temporal patterns in how professional coaching provided by BetterUp influences a change in various positive mental health and psychological well-being outcomes. This study provides a novel look at how behavioral change can evolve over time with support and scaffolding provided by a professional coach. To our knowledge, this is the first study to use multiple assessments to capture the change within individuals from before starting coaching to midway through their interaction and after approximately 6 months of being engaged in coaching. The multiple repeated assessments allowed us to analyze and understand a timeline of what happens throughout the longitudinal coaching journey.

The results reported herein can be used as a guide for coaches to stage the type of support they provide to their clients, which can enable early mental and behavioral changes that support improvement in skills that develop more slowly. Additionally, we identified professional coaching as a resource for employees that escapes the stigma associated with more traditional employee assistance and well-being programs. Professional one-on-one coaching empowers individuals to tackle a range of negative to positive mental health states with ongoing, personalized support that leaves them best prepared to excel personally and professionally, while contributing to the broader success of their organization.

Conflicts of Interest

The authors are employed by and receive income from BetterUp, Inc, and/or have been granted equity in BetterUp, Inc.

Multimedia Appendix 1
Scale development technical report.
[PDF File (Adobe PDF File), 816 KB - jmir_v23i7e27774_app1.pdf]

Multimedia Appendix 2
R script for reported analyses.
[PDF File (Adobe PDF File), 41 KB - jmir_v23i7e27774_app2.pdf]

References

1. Kortum E, Leka S, Cox T. Psychosocial risks and work-related stress in developing countries: health impact, priorities, barriers and solutions. Int J Occup Med Environ Health 2010;23(3):225-238. [doi: 10.2478/v10001-010-0024-5] [Medline: 20934955]
2. Peacock A, Leung J, Larney S, Colledge S, Hickman M, Rehm J, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction 2018 Oct;113(10):1905-1926. [doi: 10.1111/add.14234] [Medline: 29749059]
3. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
4. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
5. Besteher B, Gaser C, Nenadić I. Brain structure and subclinical symptoms: a dimensional perspective of psychopathology in the depression and anxiety spectrum. Neuropsychobiology 2020;79(4-5):270-283 [FREE Full text] [doi: 10.1159/000501024] [Medline: 31340207]
6. Cukrowicz KC, Schlegel EF, Smith PN, Jacobs MP, Van Orden KA, Paukert AL, et al. Suicide ideation among college students evidencing subclinical depression. J Am Coll Health 2011;59(7):575-581 [FREE Full text] [doi: 10.1080/07448481.2010.483710] [Medline: 21823951]
7. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch Gen Psychiatry 2005 Jan;62(1):66-72. [doi: 10.1001/archpsyc.62.1.66] [Medline: 15630074]
8. Gould CE, Karna R, Jordan J, Kawai M, Hirst R, Hantke N, et al. Subjective but not objective sleep is associated with subsyndromal anxiety and depression in community-dwelling older adults. Am J Geriatr Psychiatry 2018 Jul;26(7):806-811 [FREE Full text] [doi: 10.1016/j.jagp.2018.03.010] [Medline: 29709510]

XSL•FO
RenderX

9. Zhu Y, Qi S, Zhang B, He D, Teng Y, Hu J, et al. Connectome-based biomarkers predict subclinical depression and identify abnormal brain connections with the lateral habenula and thalamus. Front Psychiatry 2019;10:371. [doi: 10.3389/fpsyt.2019.00371] [Medline: 31244688]
10. Friedman EM, Ryff CD. Living well with medical comorbidities: a biopsychosocial perspective. J Gerontol B Psychol Sci Soc Sci 2012 Sep;67(5):535-544 [FREE Full text] [doi: 10.1093/geronb/gbr152] [Medline: 22377799]
11. Keyes CLM, Dhingra SS, Simoes EJ. Change in level of positive mental health as a predictor of future risk of mental illness. Am J Public Health 2010 Dec;100(12):2366-2371. [doi: 10.2105/AJPH.2010.192245] [Medline: 20966364]
12. Montpetit MA, Bergeman CS, Bisconti TL, Rausch JR. Adaptive change in self-concept and well-being during conjugal loss in later life. Int J Aging Hum Dev 2006;63(3):217-239. [doi: 10.2190/86WW-652A-M314-4YLA] [Medline: 17152410]
13. Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166. [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]
14. Sheldon KM, Lyubomirsky S. Achieving sustainable gains in happiness: change your actions, not Your circumstances*. J Happiness Stud 2006 Mar;7(1):55-86. [doi: 10.1007/s10902-005-0868-8]
15. Hone LC, Jarden A, Duncan S, Schofield GM. Flourishing in New Zealand workers: associations with lifestyle behaviors, physical health, psychosocial, and work-related indicators. J Occup Environ Med 2015 Sep;57(9):973-983. [doi: 10.1097/JOM.0000000000000508] [Medline: 26340286]
16. Boehm JK, Lyubomirsky S. Does happiness promote career success? J Career Assess 2008 Feb 01;16(1):101-116. [doi: 10.1177/1069072707308140]
17. Attridge M. Employee assistance programsvidencecurrent trends. In: Gatchel RJ, Schultz LZ, editors. Handbook of Occupational Health and Wellness. Cham: Springer; 2012.
18. Attridge M. A global perspective on promoting workplace mental health and the role of employee assistance programs. Am J Health Promot 2019 May;33(4):622-629. [doi: 10.1177/0890117119838101c] [Medline: 31006254]
19. Attridge M, Cahill T, Granberry SW, Herlihy PA. The national behavioral consortium industry profile of external EAP vendors. J Workplace Behav Health 2013 Oct;28(4):251-324. [doi: 10.1080/15555240.2013.845050]
20. Mattke S, Schnyer C, Van BK. A review of the U.S. workplace wellness market. RAND Corporation. 2012. URL: https://www.rand.org/pubs/occasional_papers/OP373.html [accessed 2020-11-15]
21. Pollitz K, Rae M. Workplace Wellness Programs Characteristics and Requirements. Kaiser Family Foundation. 2016. URL: https://www.kff.org/private-insurance/issue-brief/workplace-wellness-programs-characteristics-and-requirements/ [accessed 2021-06-19]
22. Song Z, Baicker K. Effect of a workplace wellness program on employee health and economic outcomes: a randomized clinical trial. JAMA 2019 Apr 16;321(15):1491-1501 [FREE Full text] [doi: 10.1001/jama.2019.3307] [Medline: 30990549]
23. Eddington D, Pitts J. Shared Values, Shared Results: Positive Organizational Health as a Win-Win Philosophy. Washington, DC: Edington Associates; 2015.
24. Goetzel RZ, Henke RM, Tabrizi M, Pelletier KR, Loeppke R, Ballard DW, et al. Do workplace health promotion (wellness) programs work? J Occup Environ Med 2014 Sep;56(9):927-934. [doi: 10.1097/JOM.0000000000000276] [Medline: 25153303]
25. Joseph B, Walker A, Fuller-Tyszkiewicz M. Evaluating the effectiveness of employee assistance programmes: a systematic review. Eur J Work Organiz Psychol 2017 Sep 06;27(1):1-15. [doi: 10.1080/1359432x.2017.1374245]
26. Kelloway EK. Mental health in the workplace: Towards evidence-based practice. Can Psychol 2017;58(1):1-6. [doi: 10.1037/cap0000084]
27. Kirk AK, Brown DF. Employee assistance programs: a review of the management of stress and wellbeing through workplace counselling and consulting. Austral Psychol 2006 Aug 16;38(2):138-143. [doi: 10.1080/00050060310001707137]
28. Jones RJ, Woods SA, Guillaume YRF. The effectiveness of workplace coaching: A meta-analysis of learning and performance outcomes from coaching. J Occup Organ Psychol 2015 Apr 15;89(2):249-277. [doi: 10.1111/joop.12119]
29. Theeboom T, Beersma B, Van VA. Coaching in Organizations? A Meta-Analytic Review of Individual Level Effects. In: Academy of Management Proceedings. Briarcliff Manor, NY: Academy of Management; 2015.
30. Vanhove AJ, Herian MN, Perez ALU, Harms PD, Lester PB. Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. J Occup Organ Psychol 2015 Apr 25;89(2):278-307. [doi: 10.1111/joop.12123]
31. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Person Soc Psychol 1989;57(6):1069-1081. [doi: 10.1037/0022-3514.57.6.1069]
32. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
33. Bonanno GA. Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? Am Psychol 2004 Jan;59(1):20-28. [doi: 10.1037/0003-066X.59.1.20] [Medline: 14736317]
34. Diener E. Subjective well-being. Psychol Bull 1984 May;95(3):542-575. [Medline: 6399758]
35. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]

XSL•FO
RenderX

36. Ellinger AD, Ellinger AE, Keller SB. Supervisory coaching behavior, employee satisfaction, and warehouse employee performance: A dyadic perspective in the distribution industry. Hum Resource Dev Quart 2003 Nov 24;14(4):435-458. [doi: 10.1002/hrdq.1078]
37. Hamlin RG, Ellinger AD, Beattie RS. Coaching at the heart of managerial effectiveness: A cross-cultural study of managerial behaviours. Hum Resource Dev Int 2006 Sep;9(3):305-331. [doi: 10.1080/13678860600893524]
38. Ting S, Hart E. Formal coaching. In: The Center for Creative Leadership handbook of leadership development. Greensboro: The Center for Creative Leadership; 2004:116-150.
39. Salas E, Kozlowski S. Learning, training, and development in organizations: Much progress and a peek over the horizon. London: Taylor and Francis; 2010.
40. Grant A. Towards a psychology of coaching: the impact of coaching on metacognition, mental health and goal attainment. Sydney: Macquarie University; 2001.
41. Core Competencies. International Coaching Federation. 2020. URL: https://coachfederation.org/core-competencies [accessed 2020-12-04]
42. Ryff CD, Singer B. Interpersonal flourishing: a positive health agenda for the new millennium. Pers Soc Psychol Rev 2016 Dec 21;4(1):30-44. [doi: 10.1207/s15327957pspr0401_4]
43. Ryff CD, Keyes CLM. The structure of psychological well-being revisited. J Person Soc Psychol 1995;69(4):719-727. [doi: 10.1037/0022-3514.69.4.719]
44. Morozink JA, Friedman EM, Coe CL, Ryff CD. Socioeconomic and psychosocial predictors of interleukin-6 in the MIDUS national sample. Health Psychol 2010 Nov;29(6):626-635 [FREE Full text] [doi: 10.1037/a0021360] [Medline: 20954777]
45. Hutchinson D, Waters S, Yost A, Sinar E, Kellerman GR. Whole Person Model 2.0 Technical Documentation. San Francisco, CA: Available upon request to BetterUp Labs, info@betterup.co; Attn: BetterUp Labs; 2019.
46. Avolio BJ, Gardner WL. Authentic leadership development: getting to the root of positive forms of leadership. Leadership Quart 2005 Jun;16(3):315-338. [doi: 10.1016/j.leaqua.2005.03.001]
47. Luthans F, Luthans KW, Luthans BC. Positive psychological capital: beyond human and social capital. Business Horizons 2004 Jan;47(1):45-50. [doi: 10.1016/j.bushor.2003.11.007]
48. Seligman MEP. Flourish: A visionary new understanding of happiness and well-being. New York, NY: Simon & Schuster; 2011.
49. Luszczynska A, Diehl M, Gutiérrez-Doña B, Kuusinen P, Schwarzer R. Measuring one component of dispositional self-regulation: attention control in goal pursuit. Person Ind Diff 2004 Aug;37(3):555-566. [doi: 10.1016/j.paid.2003.09.026]
50. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb 10;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
51. Lips-Wiersma M, Wright S. Measuring the meaning of meaningful work. Group Organiz Manag 2012 Oct 10;37(5):655-685. [doi: 10.1177/1059601112461578]
52. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
53. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]
54. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
55. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
56. Kuznetsova A, Brockhoff PB, Christensen RHB. LMER test package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
57. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 2017 Sep;14(134):20170213 [FREE Full text] [doi: 10.1098/rsif.2017.0213] [Medline: 28904005]
58. Cumming G. The new statistics: why and how. Psychol Sci 2014 Jan;25(1):7-29. [doi: 10.1177/0956797613504966] [Medline: 24220629]
59. Huppert F. Positive mental health in individuals and populations. In: Huppert FA, Baylis N, Keverne B, editors. The science of well-being. Oxford, UK: Oxford University Press; 2005:307.
60. Gilbert DT, Wilson TD. Prospection: experiencing the future. Science 2007 Sep 07;317(5843):1351-1354. [doi: 10.1126/science.1144161] [Medline: 17823345]
61. Seligman MEP, Railton P, Baumeister RF, Sripada C. Navigating into the future or driven by the past. Perspect Psychol Sci 2013 Mar;8(2):119-141. [doi: 10.1177/1745691612474317] [Medline: 26172493]
62. Baumeister RF, Vohs KD, Oettingen G. Pragmatic prospection: how and why people think about the future. Rev Gen Psychol 2016 Mar 01;20(1):3-16. [doi: 10.1037/gpr0000060]

XSL•FO
RenderX

63. Tugade MM, Fredrickson BL. Regulation of positive emotions: emotion regulation strategies that promote resilience. J Happiness Stud 2006 Dec 19;8(3):311-333. [doi: 10.1007/s10902-006-9015-4]
64. Bliese PD, Edwards JR, Sonnentag S. Stress and well-being at work: a century of empirical trends reflecting theoretical and societal influences. J Appl Psychol 2017 Mar;102(3):389-402. [doi: 10.1037/apl0000109] [Medline: 28125263]
65. House J. The nature of social support. In: Work stress and social support. Boston, MA: Addison-Wesley Publishing Company; 1981:13-30.
66. Steger MF, Dik BJ, Duffy RD. Measuring meaningful work. J Career Assess 2012 Feb 19;20(3):322-337. [doi: 10.1177/1069072711436160]

Abbreviations

ANOVA: analysis of variance
PERMA: positive emotion, engagement, relationships, meaning, accomplishments
PHQ-4: Patient Health Questionnaire 4

Edited by R Kukafka; submitted 08.02.21; peer-reviewed by C Gould, J Zhang; comments to author 25.02.21; revised version received 11.03.21; accepted 13.05.21; published 05.07.21.

Please cite as:
Jeannotte AM, Hutchinson DM, Kellerman GR
Time to Change for Mental Health and Well-being via Virtual Professional Coaching: Longitudinal Observational Study
J Med Internet Res 2021;23(7):e27774
URL: https://www.jmir.org/2021/7/e27774
doi:10.2196/27774
PMID:33993102

©Alexis M Jeannotte, Derek M Hutchinson, Gabriella R Kellerman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy of a Just-in-Time Adaptive Intervention to Promote HIV Risk Reduction Behaviors Among Young Adults Experiencing Homelessness: Pilot Randomized Controlled Trial

Diane Santa Maria[1], DrPH, MSN, BSN; Nikhil Padhye[1], PhD; Michael Businelle[2], PhD; Yijiong Yang[1], MHA; Jennifer Jones[1], MPH; Alexis Sims[1], MPH; Marguerita Lightfoot[3], PhD

[1]Cizik School of Nursing, University of Texas Health Science Center at Houston, Houston, TX, United States

[2]TSET Health Promotion Research Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[3]Center for AIDS Prevention Studies and UCSF Prevention Research Center, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Diane Santa Maria, DrPH, MSN, BSN
Cizik School of Nursing
University of Texas Health Science Center at Houston
6901 Bertner Ave
Houston, TX, 77030
United States
Phone: 1 713 500 2190
Email: diane.m.santamaria@uth.tmc.edu

Abstract

Background: People experiencing homelessness have higher rates of HIV than those who are stably housed. Mental health needs, substance use problems, and issues unique to homelessness such as lack of shelter and transiency need to be considered with regard to HIV prevention. To date, HIV prevention interventions for young adults experiencing homelessness have not specifically addressed modifiable real-time factors such as stress, sexual or drug use urge, or substance use, or been delivered at the time of heightened risk. Real-time, personalized HIV prevention messages may reduce HIV risk behaviors.

Objective: This pilot study tested the initial efficacy of an innovative, smartphone-based, just-in-time adaptive intervention that assessed predictors of HIV risk behaviors in real time and automatically provided behavioral feedback and goal attainment information.

Methods: A randomized attention control design was used among young adults experiencing homelessness, aged 18-25 years, recruited from shelters and drop-in centers in May 2019. Participants were randomized to either a control or an intervention group. The intervention (called MY-RID [Motivating Youth to Reduce Infection and Disconnection]) consisted of brief messages delivered via smartphone over 6 weeks in response to preidentified predictors that were assessed using ecological momentary assessments. Bayesian hierarchical regression models were used to assess intervention effects on sexual activity, drug use, alcohol use, and their corresponding urges.

Results: Participants (N=97) were predominantly youth (mean age 21.2, SD 2.1 years) who identified as heterosexual (n=51, 52%), male (n=56, 57%), and African American (n=56, 57%). Reports of sexual activity, drug use, alcohol use, stress, and all urges (ie, sexual, drug, alcohol) reduced over time in both groups. Daily drug use reduced by a factor of 13.8 times over 6 weeks in the intervention group relative to the control group (Multimedia Appendix 4). Lower urges for sex were found in the intervention group relative to the control group over the duration of the study. Finally, there was a statistically significant reduction in reports of feeling stressed the day before between the intervention and control conditions (P=.03).

Conclusions: Findings indicate promising intervention effects on drug use, stress, and urges for sex in a hard-to-reach, high-risk population. The MY-RID intervention should be further tested in a larger randomized controlled trial to further investigate its efficacy and impact on sexual risk behaviors.

Trial Registration: ClinicalTrials.gov NCT03911024; https://clinicaltrials.gov/ct2/show/NCT03911024

(J Med Internet Res 2021;23(7):e26704) doi:10.2196/26704

XSL•FO
RenderX

KEYWORDS

youth; homelessness; HIV; prevention; just-in-time adaptive interventions; mHealth; ecological momentary assessments; drug use; stress; intervention; smartphone; mobile phone; drug; efficacy; pilot; feasibility; predictor; risk; behavior

Introduction

Background

Young adult homelessness continues to be a major public health problem with 1 in 10 young adults aged 18 to 25 years experiencing homelessness over the course of a year [1] and an estimated 1.7 to 2.5 million youth under 25 years experiencing homelessness each year in the United States [2]. Securing food and shelter while experiencing the hardships and dangers of living on the streets creates enormous challenges to maintaining one's health and well-being [3]. The mortality rate for youth experiencing homelessness (YEH) is 5 to 10 times higher than peers in the general population [4], and many YEH have chronic mental and physical conditions, engage in substance use, and have unmet health and mental health care needs [3,5,6]. One persistent health concern is HIV. Unstable housing is a significant barrier to accessing and engaging in HIV care, maintaining viral suppression, and reducing HIV transmission [7]. The implications of mental health needs, substance use problems, and issues unique to YEH such as the lack of stable sheltering options need to be considered with regard to HIV prevention. This is particularly salient as people experiencing homelessness have higher rates of HIV than those who are stably housed [8]. While HIV prevalence data for YEH are sparse, one study found a self-reported HIV diagnoses rate of 4% [9].

HIV Risk Among Youth

HIV risk among all youth is correlated with sexual orientation [10], childhood abuse [11,12], and histories of foster/juvenile justice involvement [13-16]. Further, modifiable factors have been found to predict HIV risk in nonhomeless youth populations. Stress, sexual urge, and substance use negatively impact sexual decision-making and increase HIV risk [17-19]. HIV risk behaviors such as condomless sex and substance use are also correlated with factors such as stress [20,21] and depression [22]. Stress has also been correlated with inconsistent condom use; number of sexual partners; and substance use in young females [23-27], African American adolescent females [28], urban Black heterosexual men [29], and young men who have sex with men (MSM) [30-32]. Finally, a history of traumatic stress and current substance use are associated with more frequent sexual urges among MSM [33].

Young adults engage in HIV risk behaviors [17-19]. Engagement in risk behaviors may be heightened by low motivation for HIV prevention related to time spent on the streets [18] and exacerbated by high levels of trauma experienced prior to and while homeless [34,35]. In a systematic review of sexual behaviors among YEH, most studies did not examine how situational variables affect sexual risk [34] despite the mounting evidence of the significant correlations between elevated stress, sexual urge, and substance use with HIV risk behaviors in female youth [26,27,36,37] and YEH [38,39]. For example, experiencing sexual urges has been found to influence YEH's decision to engage in condomless sex [40]. Substance use is also associated with condomless sex and sexual victimization among homeless and urban youth [35,41,42]. Finally, in one study using ecological momentary assessments (EMAs), the odds of having sex on a given day were found to be highest on days when YEH experienced sexual urge and used drugs, and the odds of substance use were highest on the days when youth experienced high stress and drug urge [43].

HIV Prevention Among YEH

To date, HIV prevention interventions for YEH have not specifically addressed modifiable real-time factors such as stress, sexual urge, or substance use, or been delivered at the time of heightened risk. Using real-time, personalized HIV prevention messages may provide more timely information and produce more motivation for behavioral change than those seen in prior interventions. Importantly, when programs are tailored and relevant, YEH are interested in health promotion programs, can be recruited and retained in interventions and research studies [44,45], and demonstrate improved outcomes despite challenges with sustaining intervention engagement. In a group-based study with YEH, there were positive intervention effects on alcohol use, motivation, and condom self-efficacy even though 52% of participants did not attend all intervention sessions [46].

One method to reach and engage YEH in interventions is to utilize mobile technology. Interventions that provide personal motivational messages in response to real-time thoughts, feelings, sexual urges, and substance use may be more effective than interventions that are homogeneous and primarily informational in nature [47,48]. Interventions that can be delivered via smartphones at the time of heightened HIV risk may place tailored health messages more proximally to critical behavioral decision points, thereby increasing access to and relevance of prevention interventions [49].

Assessing Real-Time HIV Risk

EMA is currently the gold standard and most accurate way to measure real-time factors in natural settings [50,51], with high compliance rates (78%) found among youth across 42 studies [52]. EMA assesses within-person variance to risk exposures (eg, where, when, and with whom sexual risk is likely to occur throughout a day) by capturing repeated measures to assess changes in behaviors, cognition, environmental factors, and symptoms. Therefore, EMA has the potential for use as a driver of interventions that tailors messages to one's current risk level. Affect and behavioral monitoring associated with EMA may also increase self-awareness of HIV risk by capturing behavioral patterns and assessing predictors of risk, including sexual urge, stress, and substance use [53]. In smoking studies, EMA alone has lowered urges and stress [54]. By increasing one's awareness of risk, EMA can detect risk before the behavior occurs. EMA data that are collected at or near the moment when HIV risk behaviors occur can reduce recall biases that are associated with other measures. For example, one study found that 54% of youth reported condom use during their last sexual encounter at baseline, yet 76% of sex acts were condomless when assessed

XSL•FO
RenderX

in real time using EMA [46]. Recall data have higher potential for bias, neglect intraindividual variability, and do not capture risk and protective factors as they occur in the real world. Consistently high EMA completion rates have also been found among youth, including substance use (80%) [55], substance use recovery (87%) [56], smoking (88%) [57], sexual behaviors (80%) [58], and drinking (89%) [59]. Little is known about EMA compliance rates among YEH, although one study found that 89% of participants provided EMA data during the study with a 62% compliance rate [60].

Delivering Interventions in Response to Real-Time Risk

Just-in-time adaptive interventions (JITAIs), such as EMA-driven messages, may be an effective delivery strategy for information and motivational messages to be delivered at the time of heightened risk detection prior to engaging in a risk behavior and in response to a reported risk behavior antecedent. JITAIs can target the proximal, modifiable mediators that indicate the emergence of a vulnerable state (eg, high sexual urge, substance use, or spikes in stress). For example, messages can be delivered in response to elevated sexual urge and/or after a recent sexual assault [43,60-62]. JITAIs can be effectively delivered for a variety of health behaviors and psychological symptoms management [63], and real-time messaging has improved risk behaviors, including long-term smoking cessation [64], binge drinking among young adults with hazardous alcohol use [65], and sexual risk behaviors and sexually transmitted infection (STI) testing among youth [66]. JITAIs have been found to significantly reduce anxiety and stress [67-69], alcohol use [70], and depressive symptoms [71], and increase pre-exposure prophylaxis (PrEP) uptake among MSM [72]. EMA-informed JITAIs build off of the willingness to disclose personal information electronically through EMAs [73], overcome geographic and organizational barriers to reaching the underserved [74], require few agency resources, are easily accessible to youth, address personalized prevention care, and are particularly attractive to young people especially when these interventions are developed with the target audience to enhance sustainable use [75].

Optimizing the interactivity that smartphones provide, it is possible to combine tested mobile health (mHealth) strategies (eg, text messaging) [76] with other proven technology-based strategies such as tailored education [77,78] and motivational messaging [79,80] to deliver scalable, cost-efficient HIV prevention interventions. These strategies have had positive outcomes for smoking cessation in youth [81] and among homeless adults [82,83], treatment adherence among youth living with HIV [78], and HIV prevention in African American youth [84,85]. Such real-time interventions may address the challenges of reaching YEH related to transiency and heterogeneity [86] by targeting real-time factors such as sexual urge, substance use, and stress at the time of heightened HIV risk [34]. JITAIs can deliver personalized HIV prevention messages that vary in content and dose depending on an individual's current sexual urges, substance use, and spikes in stress [87], providing the right type and dose at the optimal time [88]. To our knowledge, no JITAIs have been developed for YEH that use EMA and deliver personalized, time-varied HIV prevention messages. This study advances work done in this field by evaluating the preliminary efficacy of an EMA-driven, personalized HIV prevention intervention that is sensitive to variability in risk among a sample of YEH.

Purpose of This Study

No interventions to date have been carried out to intervene at the individual level to increase HIV risk perception and behavioral self-monitoring for the prevention of drug use and sexual risk behaviors among YEH. Therefore, in this study we leveraged the use of mobile technology to test the initial efficacy of a beta-tested mobile, just-in-time, adaptive HIV prevention messaging intervention called MY-RID (Motivating Youth to Reduce Infection and Disconnection; pronounced "My Ride") among YEH on sexual behaviors and substance use. MY-RID is an innovative, theory-based (Information, Motivation, and Behavioral Skills [IMB] model [89]) JITAI, delivered via a smartphone app, that targets real-time predictors of HIV risk behaviors at the time of high risk, responds to patterns of risk behaviors to motivate participants to engage in prevention services, and provides behavioral feedback and goal attainment information in real time [88].

Methods

Study Design

The study used a 1:1 randomized attention control design to pilot test the initial efficacy of MY-RID. All study protocols were approved by the Institutional Review Board for the Protection of Human Subjects prior to recruitment. Interested youth were screened, consented, and then completed a baseline survey (Multimedia Appendix 1) about demographic characteristics that was administered on an iPad. Participants were then randomized to either the control or the intervention group, and every participant received an Android smartphone with unlimited data. The phone was preloaded with the MY-RID app, which uses the INSIGHT mHealth platform [90]. The MY-RID app prompted and delivered once daily EMAs and targeted messages (ie, intervention messages or control condition messages) in response to EMA data for 6 weeks. All participants were oriented to the EMA procedures and indicated the time they typically woke up and went to bed each day to assure EMA and messages did not disturb their sleep.

Recruitment

YEH aged 18-25 years were recruited from a shelter and drop-in center serving YEH over the course of 1 week in May 2019 from the Houston, TX, region using group presentations and flyers. Youth were eligible if they were between 18 and 25 years and were experiencing homelessness. Experiencing homelessness was defined as: (1) living on the streets or in a place not meant for human habitation, a shelter, hotel/motel, or any place they cannot stay for more than 30 days or (2) currently receiving homelessness services. In order to assure participants could complete the EMA surveys and understand the tailored intervention messages independently, participants were required to be able to speak and read English (determined by the Rapid Estimate of Adult Literacy in Medicine-Short Form [REALM]) at a 7th-grade level or higher [91].

XSL•FO
RenderX

Ecological Momentary Assessments

Participants were prompted 3 times a day to complete EMAs during the first 2 weeks of the study, two daily EMAs for the next 2 weeks, and one daily EMA for the final 2 weeks. These EMAs took about 1 to 2 minutes to complete. Example EMA items included: "I am feeling a strong urge to: (have sex, do drugs, drink alcohol)," "Select all of the drugs you used in the last 2 hours," "Did you have sex yesterday?" EMAs were prompted on the phone 30 minutes after the participants indicated wake time and took approximately 1 to 5 minutes to complete. All participants were able to receive up to $120 based on their EMA response rate (>90%=$40, 75%-89%=$35, 50%-74%=$30, 25%-49%=$20, and >24%=$15) over the 6-week period. A compensation monitor was programmed in the app to show the participants their response and compensation.

Intervention Messages

Using the IMB model to assure the messaging reflected on the cognitive, behavioral, and environmental factors of HIV [89], over 360 messages were developed and beta tested with YEH that addressed the real-time predictors of HIV risk, including urge to use drugs, sexual urge, stress, and drug use. First, the research team developed the theory-driven messages. Next, messages were presented to YEH, who offered ways they would "say to this to a friend" to increase the linguistic relevance of the messages. Finally, messages were beta tested with YEH until there was agreement that the message was likely to impact thoughts and behaviors. Example messages included: "Avoiding drugs can help reduce mental health issues." and "It's easy to catch an STD. Be careful and always use a condom." Consistent with the IMB model, messages included information and knowledge about the behavior in question (eg, substance use, sexual risk behavior), the individual's motivation to perform the behavior, and the behavioral skills necessary to perform the behavior [92].

Intervention Arm

During the baseline visit, youth randomized to the intervention arm were asked to set an HIV prevention behavior goal after a study team member reviewed basic HIV prevention strategies. Goal options included increasing condom use, decreasing the number of sexual partners, using PrEP daily, getting tested for HIV, reducing drug and alcohol use, not having sex while using drugs or drinking alcohol, and avoiding injection drug use. Data entered by the participant during the daily EMAs populated a graphic goal interface that depicted the current level of goal attainment based on EMA data. This interface was accessible on the phone at any time by the participant. After the completion of each EMA, participants in the intervention group received tailored messages that addressed the participant's goals: (1) safer sexual behaviors, (2) alcohol or drug use, (3) PrEP interest, and (4) HIV testing. Messages were selected using an algorithm that prompted messages based on current risk factors that targeted reduction in alcohol and drug use, promoted condom use, and provided urge management strategies (Multimedia Appendix 2). Generalized linear mixed models revealed several predictors of engagement in sexual activity that could increase HIV risk, including constant characteristics (race, sexual orientation), diagnosed conditions (psychosis, posttraumatic stress disorder), and time-varying predictors (urge for sex, drug use) [43]. Predictors of drug use were all time varying: urge for drug use, urge for alcohol use, urge to steal, viewing pornography, alcohol use, and experience of discrimination [60].

Attention Control Arm

Participants in the attention control condition were asked to set a behavioral goal related to general health behaviors. Goal options included getting 7 or more hours of sleep, eating 5 or more servings of fruits or vegetables, not using tobacco products, and exercising for at least 60 minutes a day. The control group answered the same EMA items as the intervention group; however, after the completion of each EMA, the control group received different messages related to healthy nutrition, physical activity, sleep hygiene, and prevention of tobacco use. These messages were not tailored to the EMA but sent in a random order.

Measures

The baseline and 6-week follow-up survey assessed the demographics that have been associated with sexual risk among YEH such as gender identity, age, education, sexual orientation, and adverse childhood experiences [93,94]. Baseline measures also assessed HIV/STI testing, stress [95], depression [96], and psychological distress [97], all of which have been validated previously in YEH studies. EMA measures assessed real-time HIV risk behaviors, including stress, sexual behaviors, sexual urge, and substance use. The EMA items utilized Likert scales and were developed from prior EMA studies [50,98] and tested in studies within the YEH population [43,60].

Statistical Analyses

Participant retention was evaluated to inform future JITAI studies with YEH. Kaplan-Meier survival functions were calculated using the survival package [99] in the R programming language (R Core Team) [100] to obtain the median participation time in the control and intervention groups, based on the total time spent completing activities in the app. Participant EMA completion rates were also examined over the 6-week study duration.

Baseline data consisting of participants' demographics and health status measures were compared between the control and intervention groups using Pearson chi-square tests. In instances of low cell counts, the Fisher exact test was calculated using SPSS Statistics 26 (IBM Corp). Outcomes were analyzed first on the basis of cumulative counts before using longitudinal models. The counts provided the number of participants who engaged in sexual activity, or substance use, or reported urges, or reported being stressed at least once over the 6-week study period. The differences in proportions between the control and intervention groups were assessed with Pearson chi-square tests or the Fisher exact test.

Longitudinal analysis was carried out with Bayesian hierarchical logistic regression models that assessed the intervention effects on engaging in sex, drug use, alcohol use, and their corresponding urges. The intervention effect was modeled

XSL•FO
RenderX

with fixed effects that included the interaction of intervention group and time. Random intercept and slope allowed the intervention effect to vary among participants. Stress experienced at the time of response and stress experienced by participants on the previous day were analyzed in a similar manner. Information about the prior distributions and more details about the analysis are provided in Multimedia Appendix 3. The solution was implemented with the *RStan* package [101] via code written and executed in the RStudio environment (RStudio, PBC) [102].

Sensitivity analysis was conducted to check the robustness of statistically significant intervention effects that resulted from the hierarchical regression models. The models were expected to provide robust intervention effects if data were missing at random. We considered the possibility of nonignorable missing mechanisms and used a tipping-point approach for the sensitivity analysis [103]. Multimedia Appendix 3 presents more information about the missingness mechanisms and sensitivity analysis.

Results

A total of 100 participants were enrolled, of which 3 were excluded due to lack of EMA data. Data were analyzed for the remaining 97 participants aged 18-25 years (mean 21.2, SD 2.1 years), of whom 48 (49.5%) were randomly assigned to the intervention group and 49 (50.5%) were randomly assigned to the control group. Over half of participants identified as heterosexual (n=51, 52%), male (n=56, 57%), and African American (n=56, 57%). There was noteworthy diversity in gender identity, sexual orientation, and racial identity (Table 1). Additionally, 40% (n=39) of participants had been involved with the juvenile justice system. Over 35% (n=37) of the participants rated their overall health status as excellent or very good, and 52% (n=51) of participants had been tested for HIV within the past 3 months. There were no statistically significant differences between the groups in terms of race or ethnicity, gender identity, HIV risk behaviors, or any of the other variables that were collected at baseline.

Kaplan-Meier survival curves showed that participants engaged with the app for 34.5 days (median), or about 5 of 6 possible weeks (95% CI 28.5-39.5). The length of participation was based on the range from the first to last time of responses received from each participant, and the median value was identical in the control and intervention groups. Another metric of participation was the participant response rate, which decreased over time, resulting in a median of 19 days (IQR 25) of useful data per participant. That corresponds to 46.3% of the maximum potential for responses, or 56.7% of the response potential during the median participation time. The median times for responding to EMAs were 160 seconds (IQR 98) for the daily EMA and 87 seconds (IQR 46) for the random assessments.

Frequencies of participant reported behaviors, urges (ie, sex, drug, alcohol), and stress during the EMA daily diaries (Table 2) revealed that there was a statistically significant (P=.03) difference in reports of feeling stressed the day before with a lower proportion of participants in the intervention group reporting feeling stressed the day before compared with control group participants (n=20, 41.7% vs n=31, 63.3%) across the 6 weeks of EMA implementation. Although not statistically significant, the intervention group displayed fewer transactional sex behaviors relative to the control group (n=5, 10.4% vs n=9, 18.4%). HIV risk behaviors were similar across gender identities, but current stress was higher among transgender/genderqueer participants compared to cisgender males and females (Table 3).

XSL•FO
RenderX

Table 1. Baseline characteristics by group.

Variable	EMA[a] participants (N=97)		*P* value[b]
	Intervention group (n=48), n (%)	Control group (n=49), n (%)	
Age			.62
18-21 years	28 (58.3)	31 (63.3)	
22-25 years	20 (41.7)	18 (36.7)	
Gender			.39
Male	25 (52.1)	31 (63.3)	
Female	20 (41.7)	14 (28.6)	
Transgender/genderqueer/other/missing	3 (6.3)	4 (8.2)	
Sexual orientation			.12
Gay	1 (2.1)	7 (14.3)	
Lesbian	4 (8.3)	2 (4.1)	
Straight (ie, not gay)	29 (60.4)	22 (44.9)	
Bisexual	9 (18.8)	14 (28.6)	
Asexual/pansexual/other	5 (10.4)	4 (8.2)	
Race or ethnicity			.06
White or Caucasian	0 (0)	3 (6.1)	
Black or African American	31 (64.6)	25 (51.0)	
Hispanic or Latino	7 (14.6)	2 (4.1)	
American Indian/Asian or Pacific Islander/other	5 (10.4)	9 (18.4)	
Multiracial	5 (10.4)	10 (20.4)	
Involved with the juvenile justice system	19 (39.6)	20 (40.8)	.90
Age of first homelessness			.07
Minor (<18 years)	27 (56.3)	18 (36.7)	
Adult (≥18 years)	21 (43.8)	30 (61.2)	
Missing	0 (0)	1 (2.0)	
Mental health			.68
Excellent	13 (27.1)	13 (26.5)	
Very good	6 (12.5)	5 (10.2)	
Good	10 (20.8)	16 (32.7)	
Fair	13 (27.1)	9 (18.4)	
Poor	6 (12.5)	5 (10.2)	
Missing	0 (0)	1 (2.0)	
Emotional problems in the last 7 days			.30
Never	4 (8.3)	2 (4.1)	
Rarely	7 (14.6)	7 (14.3)	
Sometimes	19 (39.6)	16 (32.7)	
Often	7 (14.6)	16 (32.7)	
Always	11 (22.9)	8 (16.3)	
Last HIV test			.88
Within the past 3 months	24 (50.0)	27 (55.1)	
More than 3 months	16 (33.3)	15 (30.6)	
Never been tested for HIV	8 (16.7)	7 (14.3)	

XSL•FO
RenderX

[a]EMA: ecological momentary assessment.

[b]Calculated using the Pearson chi-square test or the Fisher exact test for differences of proportion between the intervention and control groups.

Table 2. Frequency of participants who reported specific HIV risk factors by treatment group.

Variable	EMA[a] participants[b]		*P* value[c]
	Intervention group, n (%)	Control group, n (%)	
Had sex yesterday	33 (68.8)	29 (59.2)	.33
Used drugs yesterday	25 (52.1)	24 (49.0)	.76
Sex behaviors with drug use	18 (37.5)	15 (30.6)	.47
Condomless sex	8 (16.7)	14 (28.6)	.16
Urge to have sex	36 (75.0)	34 (69.4)	.54
Urge to use drug	25 (52.1)	30 (61.2)	.36
Traded sex yesterday	5 (10.4)	9 (18.4)	.27
Felt stressed yesterday	20 (41.7)	31 (63.3)	.03
Feel stressed now	30 (62.5)	36 (73.5)	.25

[a]EMA: ecological momentary assessment.

[b]Number of participants who engaged in the behavior at least once.

[c]Chi-square test was fit to test the significance of differences between the intervention and control groups. Significance level at $P<.05$.

Table 3. Frequency of participants who reported specific HIV risk factors by gender identity.

Variable	EMA[a] participants[b]			*P* value[c]
	Cis-male (n=56), n (%)	Cis-female (n=34), n (%)	Transgender, gender queer, or other (n=7), n (%)	
Had sex yesterday	36 (64.3)	22 (64.7)	4 (57.1)	.93
Used drugs yesterday	27 (48.2)	18 (52.9)	4 (57.1)	.85
Sex with drug use	20 (35.7)	10 (29.4)	3 (42.9)	.73
Condomless sex	12 (21.4)	9 (26.5)	1 (14.3)	.74
Urge to have sex	39 (69.6)	26 (76.5)	5 (71.4)	.78
Urge to use drug	28 (50.0)	21 (61.8)	6 (85.7)	.15
Traded sex yesterday	9 (16.1)	4 (11.8)	1 (14.3)	.85
Felt stressed yesterday	24 (42.9)	22 (64.7)	5 (71.4)	.08
Feel stressed now	32 (57.1)	28 (82.4)	6 (85.7)	.03

[a]EMA: ecological momentary assessment.

[b]Number of participants who engaged in the behavior at least once.

[c]Chi-square test was fit to test the significance of differences between the intervention and control groups. Significance level at $P<.05$.

Having sex, drug use, and alcohol use reduced over the 6-week EMA period in both groups (Figure 1 and Table 4). An intervention effect was observed for drug use. The odds ratio (OR) for the interaction of the intervention and time measured on the log scale (OR 0.62, 95% CI 0.39-0.97) implied that the odds of drug use reduced by a factor of 13.8 over 6 weeks, or by a factor of 8.5 over 3 weeks (mid-way point), in the intervention group relative to the control group. There was no intervention effect for having sex or alcohol use.

XSL•FO
RenderX

Figure 1. Daily totals of participants who reported using drugs over the 6-week study period, shown separately for the control and intervention groups of the study.

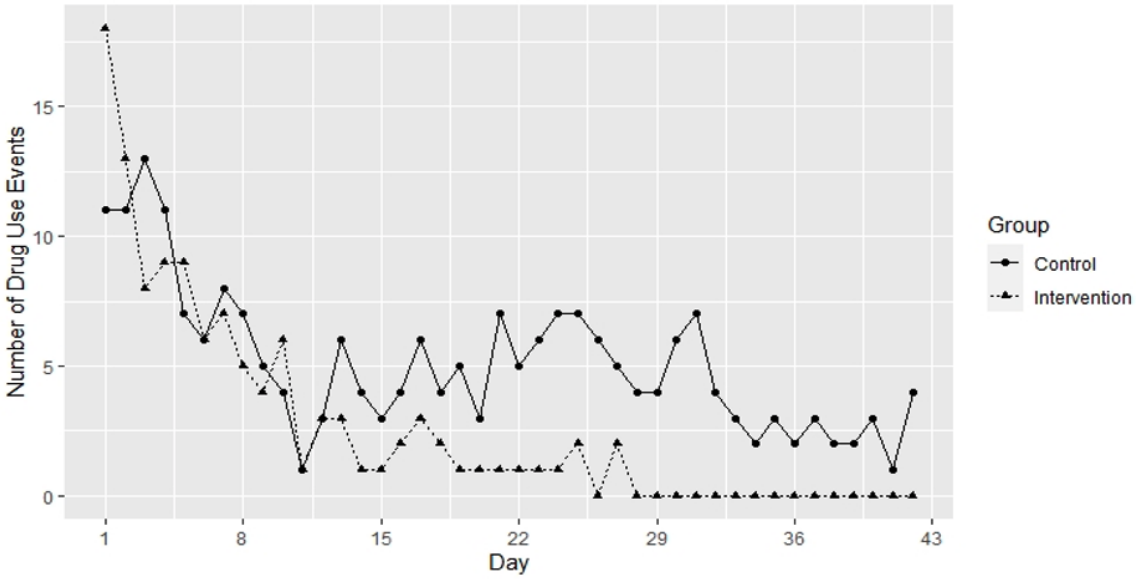


Table 4. Parameter estimates of fixed and random effects arising from Bayesian hierarchical logistic regression models for sexual intercourse, drug use, and alcohol use.

Variable	Coefficient, mean (SD)	Odds ratio (95% CI)	ESS[a]	$\hat{R}$[b]
Outcome: sexual intercourse				
Intercept	–0.863 (0.41)	0.422 (0.181-0.923)	1000	1.00
Days (log)	–0.601 (0.127)	0.548 (0.420-0.687)	563	1.00
Intervention	–0.098 (0.554)	0.907 (0.299-2.740)	1366	1.00
Intervention × days (log)	0.04 (0.158)	1.041 (0.769-1.423)	1321	1.00
Random intercept: σ	1.758 (0.285)	—[c]	1708	1.00
Random slope: σ	0.361 (0.098)	—	438	1.00
Outcome: drug use				
Intercept	–1.669 (0.663)	0.188 (0.047-0.614)	872	1.00
Days (log)	–0.597 (0.17)	0.550 (0.379-0.743)	737	1.00
Intervention	0.779 (0.861)	2.179 (0.415-12.231)	860	1.00
Intervention × days (log)	–0.486 (0.233)	0.615 (0.386-0.971)	1548	1.00
Random intercept: σ	3.212 (0.502)	—	1180	1.00
Random slope: σ	0.519 (0.138)	—	478	1.00
Outcome: alcohol use				
Intercept	–2.185 (0.541)	0.112 (0.037-0.303)	673	1.01
Days (log)	–0.564 (0.181)	0.569 (0.392-0.795)	672	1.01
Intervention	0.115 (0.652)	1.122 (0.312-4.035)	1533	1.00
Intervention × days (log)	–0.139 (0.23)	0.870 (0.545-1.344)	1603	1.00
Random intercept: σ	1.858 (0.418)	—	651	1.00
Random slope: σ	0.542 (0.14)	—	557	1.01

[a]ESS: effective sample size; after accounting for autocorrelated samples.

[b]Potential scale reduction statistic; $\hat{R}$<1.1 indicates convergence of Markov chains.

[c]Not applicable.

XSL•FO
RenderX

Urges for sex, drugs, or alcohol reduced over the study period in both groups (Table 5). Lower odds of the urge for sex were found in the intervention group (OR 0.16, 95% CI 0.04-0.63), and this was a time-independent effect that indicated suppressed sexual urge in the intervention group over the duration of the study. There was no intervention effect for drug or alcohol urge. Stress experienced now and the day before was also reduced over time in both groups (Table 6), with no intervention effect.

Table 5. Parameter estimates of fixed and random effects arising from Bayesian hierarchical logistic regression models for urges: urge for sex, urge for drug use, and urge for alcohol.

Variable	Coefficient, mean (SD)	Odds ratio (95% CI)	ESS[a]	$\hat{R}$[b]
Outcome: urge for sex				
Intercept	1.115 (0.495)	3.050 (1.157-8.004)	1757	1.00
Days (log)	–0.807 (0.145)	0.446 (0.333-0.586)	1490	1.00
Intervention	–1.845 (0.709)	0.158 (0.038-0.626)	2367	1.00
Intervention × days (log)	0.352 (0.206)	1.422 (0.947-2.143)	2141	1.00
Random intercept: σ	2.157 (0.358)	—[c]	2229	1.00
Random slope: σ	0.535 (0.105)	—	1379	1.00
Outcome: urge for drugs				
Intercept	–0.846 (0.582)	0.429 (0.130-1.296)	1427	1.00
Days (log)	–0.537 (0.171)	0.584 (0.410-0.799)	1512	1.00
Intervention	0.291 (0.809)	1.338 (0.274-6.430)	1920	1.00
Intervention × days (log)	–0.354 (0.250)	0.702 (0.428-1.148)	2499	1.00
Random intercept: σ	2.469 (0.445)	—	1283	1.00
Random slope: σ	0.611 (0.137)	—	1065	1.00
Outcome: urge for alcohol				
Intercept	–1.318 (0.709)	0.268 (0.057-0.944)	1475	1.00
Days (log)	–0.767 (0.207)	0.464 (0.297-0.679)	1272	1.00
Intervention	0.634 (0.930)	1.885 (0.313-12.884)	2500	1.00
Intervention × days (log)	–0.305 (0.294)	0.737 (0.406-1.279)	2892	1.00
Random intercept: σ	2.663 (0.544)	—	1494	1.00
Random slope: σ	0.616 (0.171)	—	772	1.01

[a]ESS: effective sample size; after accounting for autocorrelated samples.

[b]Potential scale reduction statistic; $\hat{R}$<1.1 indicates convergence of Markov chains.

[c]Not applicable.

XSL•FO
RenderX

Table 6. Parameter estimates of fixed and random effects arising from Bayesian hierarchical logistic regression models for stress experienced now and stress experienced the day before.

Variables	Coefficient, mean (SD)	Odds ratio (95% CI)	ESS[a]	$\hat{R}$[b]
Outcome: stressed now				
Intercept	1.861 (0.674)	6.430 (1.779-25.229)	2031	1.00
Days (log)	–0.762 (0.199)	0.467 (0.310-0.681)	1904	1.00
Intervention	1.448 (1.015)	4.255 (0.584-32.169)	2044	1.00
Intervention × days (log)	–0.107 (0.304)	0.899 (0.495-1.642)	2191	1.00
Random intercept: σ	3.373 (0.570)	—[c]	1623	1.00
Random slope: σ	0.902 (0.158)	—	1766	1.00
Outcome: stressed yesterday				
Intercept	–0.216 (0.418)	0.806 (0.347-1.831)	1591	1.00
Days (log)	–0.339 (0.149)	0.712 (0.526-0.942)	1445	1.00
Intervention	–0.027 (0.597)	0.973 (0.300-3.238)	2030	1.00
Intervention × days (log)	0.012 (0.213)	1.012 (0.670-1.540)	1984	1.00
Random intercept: σ	1.826 (0.417)	—	899	1.01
Random slope: σ	0.623 (0.129)	—	1050	1.01

[a]ESS: effective sample size; after accounting for autocorrelated samples.

[b]Potential scale reduction statistic; $\hat{R}$<1.1 indicates convergence of Markov chains.

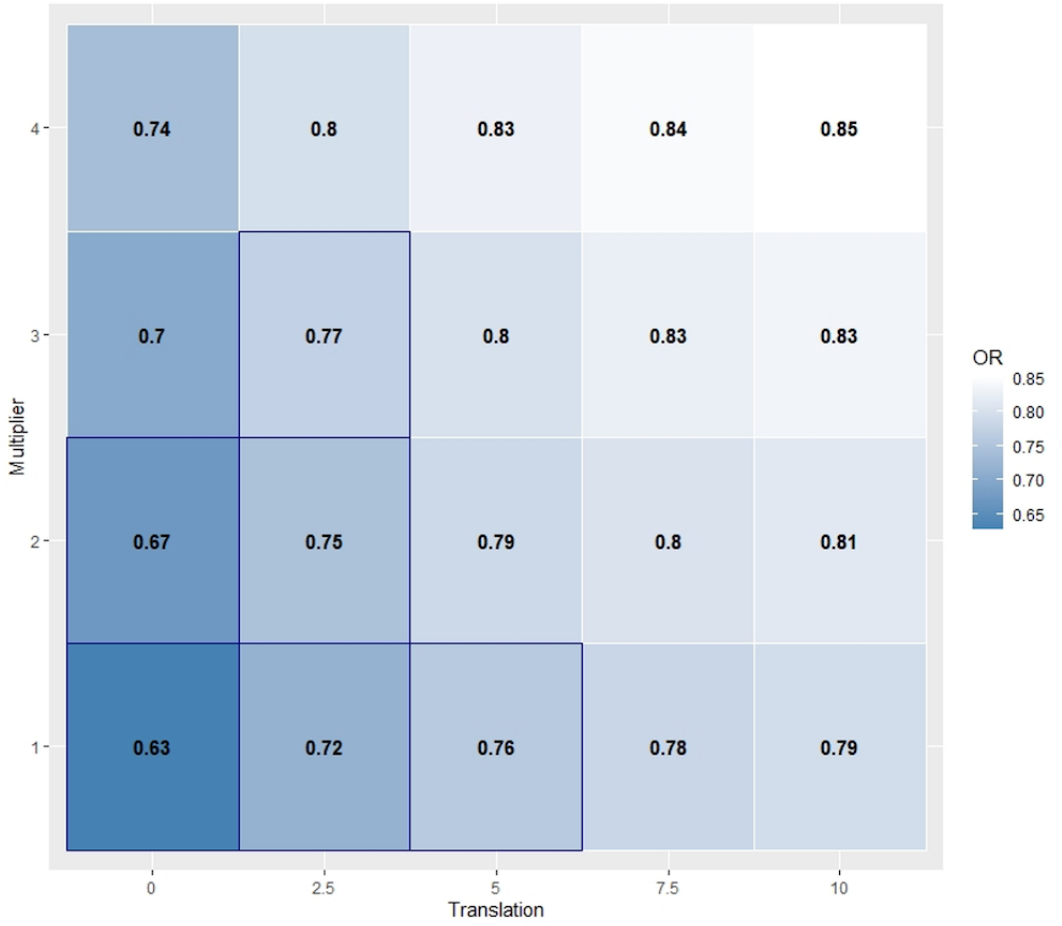
[c]Not applicable.

The number of participants with any PrEP use during the study was 17%—9 (18%) in the control group and 8 (16%) in the intervention group (χ^2_1=0.07, *P*=.79). HIV tests were completed by 11 (22%) and 13 (26%) participants in the control and intervention groups, respectively (χ^2_1=0.22, *P*=.64). Intervention goals were selected by participants at the baseline visit. Overall, participants chose from goals, including increasing condom use (n=13), reducing drug and alcohol use (n=9), reducing the number of sex partners (n=9), increasing HIV testing (n=7), using PrEP for HIV prevention (n=6), reducing injection drug use (n=4), and reducing sex while under the influence of drugs or alcohol (n=2).

Sensitivity analysis with the tipping-point approach showed that the intervention effect for drug use was fairly robust on both dimensions of the nonignorable missing mechanisms (Figure 2). The intervention effect was significant even if there was an average of two unreported drug use events among participants who never reported drug use. Similarly, the intervention effect could withstand a doubling of the probability of drug use on nonresponse days among drug users. The intervention effect could also withstand a combination of one unreported drug use event among nonusers and doubling of the probability on the other dimension. The intervention effect tipped over into nonsignificance (at the 95% confidence level) if there were two unreported drug use events simultaneously with a doubling of the probability of drug use in the second mechanism.

XSL•FO
RenderX

Figure 2. Tipping-point display of the sensitivity analysis for drug use that is based on two nonignorable missingness mechanisms. Each 2.5-unit increment in the translation levels represents roughly one unreported drug use event per nonuser during the study period. The multipliers are inflation factors of the probability of drug use on nonresponse days among those with a record of drug usage. The numbers in each cell display the odds ratio (OR) of the intervention effect, that is, the interaction of group and time. The cells with a thick navy border have $P<.05$. The OR and *P* values represent a summary of results from the multiple imputation.



Discussion

Principal Findings

This pilot randomized controlled trial of a theoretically based intervention adds to the growing body of literature on the feasibility and potential efficacy of JITAIs to address the modifiable risk factors that are correlated with HIV risk. The findings suggest promising preliminary effects of MY-RID among YEH compared with an attention control group. This intervention was effective at reducing (1) the odds of substance use, (2) the urge for sex, and (3) stress among YEH. This is the first evidence that JITAIs can reduce drug use among a high substance-using population of YEH. Given the association between substance use, as well as sexual risk behaviors (eg, condomless sex) and sexual victimization, among homeless and urban youth [35,38,39,41,42], MY-RID may be a promising approach to HIV prevention in a hard-to-reach population. These findings add to the mounting evidence of the malleability of the risk factors that impact HIV risk in real time, including stress, sexual urge, and substance use [26,27,36-39].

While this study was not powered to detect small effects or examine the impact of the intervention on condom use, the data suggest that experiencing sexual urges can negatively influence the use of condoms [40]. In addition, only 13 (27%) participants chose increasing condom use as their behavioral goal, which may also account for the lack of impact of the intervention on condomless sex. As well, while we did not examine the intervention effect on engaging in sex while using substances, the substance use literature reports that sex is highest on substance-using days [43] and that substance use is also associated with sexual risk behaviors (eg, condomless sex) and sexual victimization among youth [35,41,42].

Findings from this study further suggest that MY-RID may reduce mental health symptoms such as stress and urge, which are antecedents to HIV risks. JITAIs have been found to improve psychological symptoms [63], anxiety and stress [67-69], and depressive symptoms [71]. The evidence linking stress to sexual risk behaviors is growing [104]. Delivering coping messages when symptoms (eg, drug or alcohol use urges) are reported may improve stress management and reduce HIV risk behaviors such as substance use and risky sexual activity. Intervention effects were also found for reducing urges for sex among YEH. While reductions in sexual activity and alcohol did not reach significance, they decreased throughout the intervention period in both groups. This may indicate the need to conduct an adequately powered randomized trial to determine if these reductions in risk are significant and sustained over time.

XSL•FO
RenderX

While EMA completion rates were low among this sample of YEH (ie, EMAs were completed on 46.3% of all possible days), the average length of participation was 5 out of a total of 6 possible weeks and suggests that youth received over half of all possible intervention messages. Since intervention messages were only delivered after EMAs were completed, the average participant received messages on less than half of the study days. However, even with this low EMA completion rate and corresponding lower than expected intervention dose (less than 50%), the results indicate that the intervention demonstrated a significant impact on several key outcomes. It is important to consider that while EMA completion rates are much higher for youth EMA studies in general (78%, range 54.6%-96.2%) [52], in an EMA study with YEH, the completion rate was 62% [43]. This is the first study using EMA-responsive JITAI with YEH and, therefore, comparative data are lacking.

In a group-based study with YEH, there was a positive intervention effect on alcohol use, motivation, and condom self-efficacy with less than the prescribed intervention dose (48% completion) [46]. This suggests that a lower than planned dose may still lead to change. While comprehensive, high-dose HIV prevention interventions are effective, they also require high participant burden [105], are expensive to support, and are rarely disseminated widely. Further, while there is little evidence of the effectiveness of all interventions for YEH [106], cognitive-behavioral interventions show marginally positive results [86]. Given the difficulty in engaging YEH in HIV prevention interventions in current health care and social service delivery models, mobile, tailored JITAIs may enhance dissemination in hard-to-reach populations such as YEH. Further, since HIV risk behaviors rarely take place during engagements in care and tend to happen within one's day, JITAIs allow for interventions to be delivered more proximally at times of heightened risk. Given the intervention effect found with a lower than planned level of intervention engagement, there is a need to determine the minimal effective intervention dose to reduce participant burden.

Limitations

There are several limitations and challenges to consider when interpreting the findings of this study. First, EMA is an emerging method, and, therefore, there is no robust body of science on EMA measurement psychometrics. However, while no validated EMA scale exists for urge, a systematic review of 91 studies using EMA found the vast majority used a 1-item measure [107]. Further psychometric studies are needed on EMA measures. Second, this was a small pilot study with a convenience sample that only allowed for a preliminary examination of the intervention effects without the power to detect small effect sizes. Third, the control condition received generic motivational messages, which may have positively impacted stress management strategies, thereby diluting any effects that may be attributable to the intervention. Fourth, there may have also been reactivity to EMAs [52] similar to a Hawthorne Effect—HIV health risk behaviors went down over time in both groups, potentially because participants were thinking about their health more due to the EMAs and the generic health messages. Social desirability cannot be eliminated as a possibility that may have affected participant reports. Finally, 16 phones were reported lost or stolen during the study, one was broken, and one was returned after being found in the possession of a different participant. Despite these limitations, the strengths of the study include the diverse gender identity, sexual orientation, and race/ethnicity of the sample.

Conclusion

In conclusion, the findings from this study suggest a positive effect of a highly scalable mobile intervention that increases access to a HIV prevention intervention for a hard-to-reach population. MY-RID incorporates several of the core elements indicative of apps that have the potential to change behavior, including knowledge and information, goals and planning, feedback and monitoring, and actions [108]. This intervention builds on proven mHealth strategies [76], tailored education [77,78], and motivational messaging [79,80], delivered at the time of heightened need. Prevention interventions tailored for YEH continue to be rare and yet have led to improvements in sexual health outcomes [86,106,109-111]. Therefore, more research is needed to build on YEH's willingness to participate in intervention research studies [44,45], improvement intervention engagement, and continue to explore the mounting evidence on efficacy. For example, studies are needed to assess whether the survey length and frequency affect engagement rates. However, it is crucial to involve YEH in the development of interventions as studies suggest improved outcomes when programs are tailored and relevant [49].

Acknowledgments

Programming and technological support for the Insight app was provided through the mHealth shared resource of the Stephenson Cancer Center via an NCI Cancer Center Support Grant (grant P30CA225520). Data analysis and manuscript preparation were additionally supported by the Oklahoma Tobacco Settlement Endowment Trust (grant R21-02). The contents of the manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the sponsoring organizations.

Conflicts of Interest

MB is the primary inventor of the Insight mHealth platform, which was used to create the data collection and intervention components of the smartphone app used in this study. He receives royalties related to the use of Insight.

Multimedia Appendix 1
The MY-RID EMA survey.

XSL•FO
RenderX

[DOCX File , 23 KB - jmir_v23i7e26704_app1.docx]

Multimedia Appendix 2
Messaging algorithm.
[DOCX File , 18 KB - jmir_v23i7e26704_app2.docx]

Multimedia Appendix 3
Bayesian stats and sensitivity analysis.
[DOCX File , 22 KB - jmir_v23i7e26704_app3.docx]

Multimedia Appendix 4
Frequencies and proportions of risky behaviors.
[DOCX File , 16 KB - jmir_v23i7e26704_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 1114 KB - jmir_v23i7e26704_app5.pdf]

References

1. Morton MH, Dworsky A, Matjasko JL, Curry SR, Schlueter D, Chávez R, et al. Prevalence and Correlates of Youth Homelessness in the United States. J Adolesc Health 2018 Jan;62(1):14-21 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.10.006] [Medline: 29153445]
2. Bassuk EL, DeCandia CJ, Beach CA, Berman F. America's youngest outcasts 2014: State report card on child homelessness. National Center for Family Homelessness. Waltham, MA: American Institute for Research; 2014 Nov. URL: https://www.air.org/sites/default/files/downloads/report/Americas-Youngest-Outcasts-Child-Homelessness-Nov2014.pdf [accessed 2014-12-15]
3. Edidin JP, Ganim Z, Hunter SJ, Karnik NS. The Mental and Physical Health of Homeless Youth: A Literature Review. Child Psychiatry Hum Dev 2011 Nov 26;43(3):354-375. [doi: 10.1007/s10578-011-0270-1]
4. Baggett TP, Hwang SW, O'Connell JJ, Porneala BC, Stringfellow EJ, Orav EJ, et al. Mortality among homeless adults in Boston: shifts in causes of death over a 15-year period. JAMA Intern Med 2013 Mar 11;173(3):189-195 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1604] [Medline: 23318302]
5. Kulik D, Gaetz S, Crowe C, Ford-Jones E. Homeless youth's overwhelming health burden: A review of the literature. Paediatr Child Health 2011 Jun;16(6):e43-e47 [FREE Full text] [doi: 10.1093/pch/16.6.e43] [Medline: 22654549]
6. Hwang SW, Ueng JJM, Chiu S, Kiss A, Tolomiczenko G, Cowan L, et al. Universal Health Insurance and Health Care Access for Homeless Persons. Am J Public Health 2010 Aug;100(8):1454-1461. [doi: 10.2105/ajph.2009.182022]
7. Aidala A, Cross JE, Stall R, Harre D, Sumartojo E. Housing status and HIV risk behaviors: implications for prevention and policy. AIDS Behav 2005 Sep 9;9(3):251-265. [doi: 10.1007/s10461-005-9000-7] [Medline: 16088369]
8. Beijer U, Wolf A, Fazel S. Prevalence of tuberculosis, hepatitis C virus, and HIV in homeless people: a systematic review and meta-analysis. The Lancet Infectious Diseases 2012 Nov;12(11):859-870. [doi: 10.1016/s1473-3099(12)70177-9]
9. Santa Maria D, Flash CA, Narendorf S, Barman-Adhikari A, Petering R, Hsu H, et al. Knowledge and Attitudes About Pre-Exposure Prophylaxis Among Young Adults Experiencing Homelessness in Seven U.S. Cities. J Adolesc Health 2019 May;64(5):574-580. [doi: 10.1016/j.jadohealth.2018.06.023] [Medline: 30254009]
10. Noell JW, Ochs LM. Relationship of sexual orientation to substance use, suicidal ideation, suicide attempts, and other factors in a population of homeless adolescents. Journal of Adolescent Health 2001 Jul;29(1):31-36. [doi: 10.1016/s1054-139x(01)00205-1]
11. Melander LA, Tyler KA. The effect of early maltreatment, victimization, and partner violence on HIV risk behavior among homeless young adults. J Adolesc Health 2010 Dec;47(6):575-581. [doi: 10.1016/j.jadohealth.2010.04.010] [Medline: 21094434]
12. Tyler KA, Hoyt DR, Whitbeck LB. The Effects of Early Sexual Abuse on Later Sexual Victimization Among Female Homeless and Runaway Adolescents. J Interpers Violence 2016 Jul 02;15(3):235-250. [doi: 10.1177/088626000015003001]
13. Thompson RG, Hasin DS. Cigarette, marijuana, and alcohol use and prior drug treatment among newly homeless young adults in New York City: Relationship to a history of foster care. Drug Alcohol Depend 2011 Aug 01;117(1):66-69 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.12.020] [Medline: 21288659]
14. Fowler PJ, Toro PA, Miles BW. Pathways to and From Homelessness and Associated Psychosocial Outcomes Among Adolescents Leaving the Foster Care System. Am J Public Health 2009 Aug;99(8):1453-1458. [doi: 10.2105/ajph.2008.142547]
15. Combs KM, Begun S, Rinehart DJ, Taussig H. Pregnancy and Childbearing Among Young Adults Who Experienced Foster Care. Child Maltreat 2018 May 15;23(2):166-174 [FREE Full text] [doi: 10.1177/1077559517733816] [Medline: 29034732]

XSL•FO
RenderX

16. Zlotnick C. What research tells us about the intersecting streams of homelessness and foster care. Am J Orthopsychiatry 2009 Jul;79(3):319-325. [doi: 10.1037/a0017218] [Medline: 19839669]
17. Khawcharoenporn T, Kendrick S, Smith K. HIV risk perception and preexposure prophylaxis interest among a heterosexual population visiting a sexually transmitted infection clinic. AIDS Patient Care STDS 2012 Apr;26(4):222-233. [doi: 10.1089/apc.2011.0202] [Medline: 22404427]
18. Collins J, Slesnick N. Factors Associated with Motivation to Change HIV Risk and Substance Use Behaviors among Homeless Youth. J Soc Work Pract Addict 2011 Jul 01;11(2):163-180 [FREE Full text] [doi: 10.1080/1533256X.2011.570219] [Medline: 21765800]
19. Steinberg L. Risk Taking in Adolescence. Curr Dir Psychol Sci 2016 Jun 23;16(2):55-59. [doi: 10.1111/j.1467-8721.2007.00475.x]
20. Hill TD, Ross CE, Angel RJ. Neighborhood disorder, psychophysiological distress, and health. J Health Soc Behav 2005 Jun;46(2):170-186. [doi: 10.1177/002214650504600204] [Medline: 16028456]
21. Cutrona CE, Wallace G, Wesner KA. Neighborhood Characteristics and Depression: An Examination of Stress Processes. Curr Dir Psychol Sci 2006 Aug;15(4):188-192 [FREE Full text] [doi: 10.1111/j.1467-8721.2006.00433.x] [Medline: 18185846]
22. Diez Roux AV, Mair C. Neighborhoods and health. Ann N Y Acad Sci 2010 Mar;1186(1):125-145. [doi: 10.1111/j.1749-6632.2009.05333.x] [Medline: 20201871]
23. Solorio MR, Rosenthal D, Milburn NG, Weiss RE, Batterham PJ, Gandara M, et al. Predictors of sexual risk behaviors among newly homeless youth: a longitudinal study. J Adolesc Health 2008 Apr;42(4):401-409 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.09.023] [Medline: 18346666]
24. Wong Y, Piliavin I. Stressors, resources, and distress among homeless persons:. Social Science & Medicine 2001 Apr;52(7):1029-1042. [doi: 10.1016/S0277-9536(00)00209-4]
25. Sinha R. Chronic stress, drug use, and vulnerability to addiction. Ann N Y Acad Sci 2008 Oct;1141(1):105-130 [FREE Full text] [doi: 10.1196/annals.1441.030] [Medline: 18991954]
26. Hall KS, Moreau C, Trussell J, Barber J. Role of young women's depression and stress symptoms in their weekly use and nonuse of contraceptive methods. J Adolesc Health 2013 Aug;53(2):241-248 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.02.009] [Medline: 23582524]
27. Hall K, Kusunoki Y, Gatny H, Barber J. Stress symptoms and frequency of sexual intercourse among young women. J Sex Med 2014 Aug;11(8):1982-1990 [FREE Full text] [doi: 10.1111/jsm.12607] [Medline: 24894425]
28. Hulland E, Brown JL, Swartzendruber AL, Sales JM, Rose ES, DiClemente RJ. The association between stress, coping, and sexual risk behaviors over 24 months among African-American female adolescents. Psychol Health Med 2015;20(4):443-456 [FREE Full text] [doi: 10.1080/13548506.2014.951369] [Medline: 25159332]
29. Bowleg L, Fitz CC, Burkholder GJ, Massie JS, Wahome R, Teti M, et al. Racial discrimination and posttraumatic stress symptoms as pathways to sexual HIV risk behaviors among urban Black heterosexual men. AIDS Care 2014;26(8):1050-1057 [FREE Full text] [doi: 10.1080/09540121.2014.906548] [Medline: 24797317]
30. Parsons J, Lelutiu-Weinberger C, Botsko M, Golub SA. Predictors of day-level sexual risk for young gay and bisexual men. AIDS Behav 2013 May;17(4):1465-1477 [FREE Full text] [doi: 10.1007/s10461-012-0206-1] [Medline: 22614745]
31. Wong CF, Kipke MD, Weiss G, McDavitt B. The impact of recent stressful experiences on HIV-risk related behaviors. J Adolesc 2010 Jun;33(3):463-475 [FREE Full text] [doi: 10.1016/j.adolescence.2009.06.004] [Medline: 19608264]
32. Voisin DR, Hotton AL, Schneider JA, UConnect Study Team. The relationship between life stressors and drug and sexual behaviors among a population-based sample of young Black men who have sex with men in Chicago. AIDS Care 2017 May;29(5):545-551 [FREE Full text] [doi: 10.1080/09540121.2016.1224303] [Medline: 27590043]
33. Jerome RC, Woods WJ, Moskowitz JT, Carrico AW. The Psychological Context of Sexual Compulsivity Among Men Who Have Sex with Men. AIDS Behav 2016 Feb;20(2):273-280 [FREE Full text] [doi: 10.1007/s10461-015-1083-1] [Medline: 25957856]
34. Heerde J, Scholes-Balog KE, Hemphill SA. Associations between youth homelessness, sexual offenses, sexual victimization, and sexual risk behaviors: a systematic literature review. Arch Sex Behav 2015 Jan;44(1):181-212. [doi: 10.1007/s10508-014-0375-2] [Medline: 25411128]
35. Heerde JA, Hemphill SA. Sexual Risk Behaviors, Sexual Offenses, and Sexual Victimization Among Homeless Youth: A Systematic Review of Associations With Substance Use. Trauma Violence Abuse 2016 Dec 23;17(5):468-489. [doi: 10.1177/1524838015584371] [Medline: 25985990]
36. Hall K, Kusunoki Y, Gatny H, Barber J. Social discrimination, stress, and risk of unintended pregnancy among young women. J Adolesc Health 2015 Mar;56(3):330-337 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.11.008] [Medline: 25586228]
37. Hall K, Kusunoki Y, Gatny H, Barber J. The risk of unintended pregnancy among young women with mental health symptoms. Soc Sci Med 2014 Jan;100:62-71 [FREE Full text] [doi: 10.1016/j.socscimed.2013.10.037] [Medline: 24444840]
38. Carmona J, Slesnick N, Guo X, Letcher A. Reducing High Risk Behaviors among Street Living Youth: Outcomes of an Integrated Prevention Intervention. Child Youth Serv Rev 2014 Aug 01;43:118-123 [FREE Full text] [doi: 10.1016/j.childyouth.2014.05.015] [Medline: 25104870]

XSL•FO
RenderX

39. Castro AL, Gustafson EL, Ford AE, Edidin JP, Smith DL, Hunter SJ, et al. Psychiatric disorders, high-risk behaviors, and chronicity of episodes among predominantly African American homeless Chicago youth. J Health Care Poor Underserved 2014 Aug;25(3):1201-1216 [FREE Full text] [doi: 10.1353/hpu.2014.0124] [Medline: 25130234]
40. Kennedy D, Brown RA, Morrison P, Vie L, Ryan GW, Tucker JS. Risk evaluations and condom use decisions of homeless youth: a multi-level qualitative investigation. BMC Public Health 2015 Jan 31;15(1):62 [FREE Full text] [doi: 10.1186/s12889-015-1419-9] [Medline: 25636862]
41. Elkington K, Bauermeister JA, Zimmerman MA. Psychological distress, substance use, and HIV/STI risk behaviors among youth. J Youth Adolesc 2010 May;39(5):514-527 [FREE Full text] [doi: 10.1007/s10964-010-9524-7] [Medline: 20229264]
42. Tucker J, Ryan GW, Golinelli D, Ewing B, Wenzel SL, Kennedy DP, et al. Substance use and other risk factors for unprotected sex: results from an event-based study of homeless youth. AIDS Behav 2012 Aug;16(6):1699-1707 [FREE Full text] [doi: 10.1007/s10461-011-0017-9] [Medline: 21932093]
43. Santa Maria D, Padhye N, Yang Y, Gallardo K, Businelle M. Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study. JMIR Public Health Surveill 2018 Apr 10;4(2):e39. [doi: 10.2196/publichealth.9020]
44. Gwadz M, Rotheram-Borus MJ. Tracking high-risk adolescents longitudinally. AIDS Educ Prev 1992;Suppl:69-82. [Medline: 1389872]
45. Leonard N, Lester P, Rotheram-Borus MJ, Mattes K, Gwadz M, Ferns B. Successful recruitment and retention of participants in longitudinal behavioral research. AIDS Educ Prev 2003 Jun;15(3):269-281. [doi: 10.1521/aeap.15.4.269.23827] [Medline: 12866838]
46. Tucker JS, D'Amico EJ, Ewing BA, Miles JNV, Pedersen ER. A group-based motivational interviewing brief intervention to reduce substance use and sexual risk behavior among homeless young adults. J Subst Abuse Treat 2017 May;76:20-27 [FREE Full text] [doi: 10.1016/j.jsat.2017.02.008] [Medline: 28340904]
47. Bartholomew LK, Mullen PD. Five roles for using theory and evidence in the design and testing of behavior change interventions. J Public Health Dent 2011;71 Suppl 1:S20-S33. [doi: 10.1111/j.1752-7325.2011.00223.x] [Medline: 21656946]
48. Bartholomew Eldredge LK, Markham CM, Ruiter RAC, Fernández ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. Hoboken, NJ: John Wiley & Sons, Inc; 2016.
49. Bender K, Begun S, DePrince A, Haffejee B, Brown S, Hathaway J, et al. Mindfulness Intervention with Homeless Youth. Journal of the Society for Social Work and Research 2015 Dec;6(4):491-513. [doi: 10.1086/684107]
50. Stone A. The Science of Real-Time Data Capture: Self-Reports in Health Research. Oxford, UK: Oxford University Press; 2007.
51. Stone A, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
52. Wen C, Schneider S, Stone AA, Spruijt-Metz D. Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
53. Gaelick L, Kanfer FH, Goldstein AP. Self-management methods. In: Helping People Change: A Textbook of Methods. Oxford, UK: Pergamon Press; 1991:305-360.
54. McCarthy D, Minami H, Yeh VM, Bold KW. An experimental investigation of reactivity to ecological momentary assessment frequency among adults trying to quit smoking. Addiction 2015 Oct;110(10):1549-1560 [FREE Full text] [doi: 10.1111/add.12996] [Medline: 26011583]
55. Comulada W, Lightfoot M, Swendeman D, Grella C, Wu N. Compliance to Cell Phone-Based EMA Among Latino Youth in Outpatient Treatment. J Ethn Subst Abuse 2015;14(3):232-250 [FREE Full text] [doi: 10.1080/15332640.2014.986354] [Medline: 26114764]
56. Dennis M, Scott CK, Funk RR, Nicholson L. A Pilot Study to Examine the Feasibility and Potential Effectiveness of Using Smartphones to Provide Recovery Support for Adolescents. Subst Abus 2015;36(4):486-492 [FREE Full text] [doi: 10.1080/08897077.2014.970323] [Medline: 25310057]
57. Hoeppner B, Kahler CW, Gwaltney CJ. Relationship between momentary affect states and self-efficacy in adolescent smokers. Health Psychol 2014 Dec;33(12):1507-1517 [FREE Full text] [doi: 10.1037/hea0000075] [Medline: 25020151]
58. Shrier L, Feldman HA, Black SK, Walls C, Kendall AD, Lops C, et al. Momentary affective states surrounding sexual intercourse in depressed adolescents and young adults. Arch Sex Behav 2012 Oct;41(5):1161-1171. [doi: 10.1007/s10508-011-9787-4] [Medline: 21755382]
59. Wright C, Dietze PM, Crockett B, Lim MSC. Participatory development of MIDY (Mobile Intervention for Drinking in Young people). BMC Public Health 2016 Mar 24;16(1):184-111 [FREE Full text] [doi: 10.1186/s12889-016-2876-5] [Medline: 26911299]
60. Santa Maria D, Padhye N, Yang Y, Gallardo K, Santos GM, Jung J, et al. Drug use patterns and predictors among homeless youth: Results of an ecological momentary assessment. Am J Drug Alcohol Abuse 2018;44(5):551-560. [doi: 10.1080/00952990.2017.1407328] [Medline: 29286835]

XSL•FO
RenderX

61. Businelle M, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An Ecological Momentary Intervention for Smoking Cessation: Evaluation of Feasibility and Effectiveness. J Med Internet Res 2016 Dec 12;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
62. Businelle M, Ma P, Kendzor DE, Frank SG, Wetter DW, Vidrine DJ. Using Intensive Longitudinal Data Collected via Mobile Phone to Detect Imminent Lapse in Smokers Undergoing a Scheduled Quit Attempt. J Med Internet Res 2016 Oct 17;18(10):e275 [FREE Full text] [doi: 10.2196/jmir.6307] [Medline: 27751985]
63. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39. [doi: 10.1348/135910709X466063] [Medline: 19646331]
64. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4(4):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
65. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-72.e4 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
66. Jones K, Eathington P, Baldwin K, Sipsma H. The impact of health education transmitted via social media or text messaging on adolescent and young adult risky sexual behavior: a systematic review of the literature. Sex Transm Dis 2014 Jul;41(7):413-419. [doi: 10.1097/OLQ.0000000000000146] [Medline: 24922099]
67. Loo Gee B, Griffiths KM, Gulliver A. Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229 [FREE Full text] [doi: 10.1093/jamia/ocv043] [Medline: 25997643]
68. Grassi A, Gaggioli A, Riva G. The green valley: the use of mobile narratives for reducing stress in commuters. Cyberpsychol Behav 2009 Apr;12(2):155-161. [doi: 10.1089/cpb.2008.0156] [Medline: 19275475]
69. Riva G, Grassi A, Villani D, Gaggioli A, Preziosa A. Managing exam stress using UMTS phones: the advantage of portable audio/video support. In: Medicine Meets Virtual Reality. Milan, Italy: Applied Technology for Neuro-Psychology Lab; 2007:406-408.
70. Suffoletto B, Callaway C, Kristan J, Kraemer K, Clark DB. Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. Alcohol Clin Exp Res 2012 Mar;36(3):552-560. [doi: 10.1111/j.1530-0277.2011.01646.x] [Medline: 22168137]
71. Kauer SD, Reid SC, Crooke AHD, Khor A, Hearps SJC, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]
72. Sullivan P, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and Acceptability of a Mobile Comprehensive HIV Prevention App for Men Who Have Sex With Men: A Pilot Study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
73. Serrano K, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to Exchange Health Information via Mobile Devices: Findings From a Population-Based Survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
74. Silva B, Rodrigues JPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
75. Akter S, Ray P. mHealth - an Ultimate Platform to Serve the Unserved. Yearb Med Inform 2018 Mar 07;19(01):94-100. [doi: 10.1055/S-0038-1638697]
76. McInnes D, Li AE, Hogan TP. Opportunities for engaging low-income, vulnerable populations in health care: a systematic review of homeless persons' access to and use of information technologies. Am J Public Health 2013 Dec;103 Suppl 2(S2):e11-e24. [doi: 10.2105/AJPH.2013.301623] [Medline: 24148036]
77. Shegog R, Bartholomew LK, Parcel GS, Sockrider MM, Mâsse L, Abramson SL. Impact of a computer-assisted education program on factors related to asthma self-management behavior. J Am Med Inform Assoc 2001;8(1):49-61 [FREE Full text] [doi: 10.1136/jamia.2001.0080049] [Medline: 11141512]
78. Shegog R, Markham CM, Leonard AD, Bui TC, Paul ME. "+CLICK": pilot of a web-based training program to enhance ART adherence among HIV-positive youth. AIDS Care 2012;24(3):310-318. [doi: 10.1080/09540121.2011.608788] [Medline: 21899488]
79. Dermen KH, Thomas SN. Randomized controlled trial of brief interventions to reduce college students' drinking and risky sex. Psychol Addict Behav 2011 Dec;25(4):583-594. [doi: 10.1037/a0025472] [Medline: 21928866]
80. Baer JS, Beadnell B, Garrett SB, Hartzler B, Wells EA, Peterson PL. Adolescent change language within a brief motivational intervention and substance use outcomes. Psychol Addict Behav 2008 Dec;22(4):570-575 [FREE Full text] [doi: 10.1037/a0013022] [Medline: 19071983]
81. Whittaker R, Maddison R, McRobbie H, Bullen C, Denny S, Dorey E, et al. A multimedia mobile phone-based youth smoking cessation intervention: findings from content development and piloting studies. J Med Internet Res 2008 Nov 25;10(5):e49 [FREE Full text] [doi: 10.2196/jmir.1007] [Medline: 19033148]

XSL•FO
RenderX

82. Businelle M, Kendzor DE, Kesh A, Cuate EL, Poonawalla IB, Reitzel LR, et al. Small financial incentives increase smoking cessation in homeless smokers: a pilot study. Addict Behav 2014 Mar;39(3):717-720. [doi: 10.1016/j.addbeh.2013.11.017] [Medline: 24321696]
83. Businelle MS, Ma P, Kendzor DE, Reitzel LR, Chen M, Lam CY, et al. Predicting quit attempts among homeless smokers seeking cessation treatment: an ecological momentary assessment study. Nicotine Tob Res 2014 Oct 03;16(10):1371-1378 [FREE Full text] [doi: 10.1093/ntr/ntu088] [Medline: 24893602]
84. Cornelius JB, Cato M, Lawrence JS, Boyer CB, Lightfoot M. Development and pretesting multimedia HIV-prevention text messages for mobile cell phone delivery. J Assoc Nurses AIDS Care 2011 Sep;22(5):407-413 [FREE Full text] [doi: 10.1016/j.jana.2010.11.007] [Medline: 21256053]
85. Cornelius J, St Lawrence JS, Howard JC, Shah D, Poka A, McDonald D, et al. Adolescents' perceptions of a mobile cell phone text messaging-enhanced intervention and development of a mobile cell phone-based HIV prevention intervention. J Spec Pediatr Nurs 2012 Jan;17(1):61-69 [FREE Full text] [doi: 10.1111/j.1744-6155.2011.00308.x] [Medline: 22188273]
86. Altena AM, Brilleslijper-Kater SN, Wolf JR. Effective interventions for homeless youth: a systematic review. Am J Prev Med 2010 Jun;38(6):637-645. [doi: 10.1016/j.amepre.2010.02.017] [Medline: 20494240]
87. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. Transl Behav Med 2015 Sep 28;5(3):335-346 [FREE Full text] [doi: 10.1007/s13142-015-0324-1] [Medline: 26327939]
88. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
89. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006 Jul;25(4):462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
90. Mobile Health Technology. Health Promotion Research Center. URL: https://healthpromotionresearch.org/Mobile-Health-Technology [accessed 2020-12-12]
91. Murphy PW, Davis TC, Long SW, Jackson RH, Decker BC. Rapid estimate of adult literacy in medicine (REALM): a quick reading test for patients. Journal of Reading 1993:124-130.
92. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
93. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. American Journal of Preventive Medicine 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8]
94. Slesnick N, Zhang J, Brakenhoff B. Homeless Youths' Caretakers: The Mediating Role of Childhood Abuse on Street Victimization and Housing Instability. Soc Work 2016 Apr;61(2):147-154. [doi: 10.1093/sw/sww009] [Medline: 27180525]
95. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385-396. [doi: 10.2307/2136404]
96. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
97. Green J, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC. Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. Int J Methods Psychiatr Res 2010 Jun;19 Suppl 1(S1):23-35 [FREE Full text] [doi: 10.1002/mpr.314] [Medline: 20527003]
98. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
99. Therneau T. A Package for Survival Analysis in R, version 2. 2015. URL: https://CRAN.R-project.org/package=survival [accessed 2021-06-30]
100. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2020. URL: https://www.r-project.org/ [accessed 2021-06-30]
101. RStan: the R interface to Stan. Stan Development Team. 2016. URL: https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html [accessed 2021-06-30]
102. The R Team. RStudio: Integrated Development for R. RStudio, PBC. 2012. URL: http://www.rstudio.com/ [accessed 2021-06-30]
103. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Stat Med 2014 Oct 30;33(24):4170-4185 [FREE Full text] [doi: 10.1002/sim.6197] [Medline: 24845086]
104. Heerde J, Hemphill SA. The role of risk and protective factors in the modification of risk for sexual victimization, sexual risk behaviors, and survival sex among homeless youth: A meta-analysis. J Investig Psych Offender Profil 2016 Nov 24;14(2):150-174. [doi: 10.1002/jip.1473]
105. Arnold E, Rotheram-Borus MJ. Comparisons of prevention programs for homeless youth. Prev Sci 2009 Mar;10(1):76-86 [FREE Full text] [doi: 10.1007/s11121-008-0119-4] [Medline: 19067164]

XSL•FO
RenderX

106. Naranbhai V, Abdool Karim Q, Meyer-Weitz A. Interventions to modify sexual risk behaviours for preventing HIV in homeless youth. Cochrane Database Syst Rev 2011 Jan 19(1):CD007501 [FREE Full text] [doi: 10.1002/14651858.CD007501.pub2] [Medline: 21249691]
107. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015 Mar 01;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
108. McKay F, Slykerman S, Dunn M. The App Behavior Change Scale: Creation of a Scale to Assess the Potential of Apps to Promote Behavior Change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
109. Rotheram-Borus M, Song J, Gwadz M, Lee M, Van Rossem R, Koopman C. Reductions in HIV risk among runaway youth. Prev Sci 2003 Sep;4(3):173-187. [doi: 10.1023/a:1024697706033] [Medline: 12940468]
110. Slesnick N, Kang MJ. The impact of an integrated treatment on HIV risk behavior among homeless youth: a randomized controlled trial. J Behav Med 2008 Mar;31(1):45-59 [FREE Full text] [doi: 10.1007/s10865-007-9132-5] [Medline: 17940861]
111. Slesnick N, Prestopnik J. Dual and multiple diagnosis among substance using runaway youth. Am J Drug Alcohol Abuse 2005;31(1):179-201 [FREE Full text] [doi: 10.1081/ada-200047916] [Medline: 15768577]

Abbreviations

EMA: ecological momentary assessment
IMB: Information, Motivation, and Behavioral Skills
JITAI: just-in-time adaptive intervention
mHealth: mobile health
MSM: men who have sex with men
MY-RID: Motivating Youth to Reduce Infection and Disconnection
OR: odds ratio
PrEP: pre-exposure prophylaxis
REALM: Rapid Estimate of Adult Literacy in Medicine-Short Form
YEH: youth experiencing homelessness

Edited by G Eysenbach; submitted 22.12.20; peer-reviewed by T Lillie, J Glass, M Step; comments to author 07.01.21; revised version received 26.02.21; accepted 05.05.21; published 06.07.21.

Please cite as:
Santa Maria D, Padhye N, Businelle M, Yang Y, Jones J, Sims A, Lightfoot M
Efficacy of a Just-in-Time Adaptive Intervention to Promote HIV Risk Reduction Behaviors Among Young Adults Experiencing Homelessness: Pilot Randomized Controlled Trial
J Med Internet Res 2021;23(7):e26704
URL: https://www.jmir.org/2021/7/e26704
doi:10.2196/26704
PMID:34255679

©Diane Santa Maria, Nikhil Padhye, Michael Businelle, Yijiong Yang, Jennifer Jones, Alexis Sims, Marguerita Lightfoot. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Behavioral Activation and Depression Symptomatology: Longitudinal Assessment of Linguistic Indicators in Text-Based Therapy Sessions

Hannah A Burkhardt[1], BSc; George S Alexopoulos[2], MD; Michael D Pullmann[3], PhD; Thomas D Hull[4], PhD; Patricia A Areán[3], PhD; Trevor Cohen[1], MBChB, PhD

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States
[2]Weill Cornell Institute of Geriatric Psychiatry, Weill Cornell Medicine, White Plains, NY, United States
[3]ALACRITY Center, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States
[4]Talkspace, New York, NY, United States

Corresponding Author:
Hannah A Burkhardt, BSc
Department of Biomedical Informatics and Medical Education
University of Washington
850 Republican St
Seattle, WA, 98109
United States
Phone: 1 206 897 4636
Email: haalbu@uw.edu

Abstract

Background: Behavioral activation (BA) is rooted in the behavioral theory of depression, which states that increased exposure to meaningful, rewarding activities is a critical factor in the treatment of depression. Assessing constructs relevant to BA currently requires the administration of standardized instruments, such as the Behavioral Activation for Depression Scale (BADS), which places a burden on patients and providers, among other potential limitations. Previous work has shown that depressed and nondepressed individuals may use language differently and that automated tools can detect these differences. The increasing use of online, chat-based mental health counseling presents an unparalleled resource for automated longitudinal linguistic analysis of patients with depression, with the potential to illuminate the role of reward exposure in recovery.

Objective: This work investigated how linguistic indicators of planning and participation in enjoyable activities identified in online, text-based counseling sessions relate to depression symptomatology over time.

Methods: Using distributional semantics methods applied to a large corpus of text-based online therapy sessions, we devised a set of novel BA-related categories for the Linguistic Inquiry and Word Count (LIWC) software package. We then analyzed the language used by 10,000 patients in online therapy chat logs for indicators of activation and other depression-related markers using LIWC.

Results: Despite their conceptual and operational differences, both previously established LIWC markers of depression and our novel linguistic indicators of activation were strongly associated with depression scores (Patient Health Questionnaire [PHQ]-9) and longitudinal patient trajectories. Emotional tone; pronoun rates; words related to sadness, health, and biology; and BA-related LIWC categories appear to be complementary, explaining more of the variance in the PHQ score together than they do independently.

Conclusions: This study enables further work in automated diagnosis and assessment of depression, the refinement of BA psychotherapeutic strategies, and the development of predictive models for decision support.

(J Med Internet Res 2021;23(7):e28244) doi:10.2196/28244

KEYWORDS

natural language processing; text analysis; behavioral activation; depression; digital interventions; mental health

XSL•FO
RenderX

Introduction

Over 20% of adults in the United States have a mental illness [1]. Depression is among the most common mental health disorders: Over 19 million adults suffered major depressive episodes in 2019. Effective delivery of mental health services is a challenge for many reasons, including that individuals respond differently to therapy [2,3]. To maximize treatment benefits, mental health care providers must continually assess progress and adjust treatment plans [2]. From a research perspective, longitudinal information about known and hypothesized mechanisms of recovery is a prerequisite to the refinement of current interventions and can inform the development of new ones. Validated survey instruments exist to assess symptoms and other constructs relevant to therapy delivery and progress [4]; however, repeatedly filling out questionnaires places a burden on patients and providers, limiting the frequency with which these data can be collected. In contrast, using already available data created as part of routine care obviates the need for additional data collection. Additionally, it has been argued that subjective self-reports present potential limitations, for example due to cognitive and memory bias [5]; while careful scale design can alleviate these problems, objective, naturalistic measurements may be preferable.

Behavioral Activation and Engagement

The behavioral theory of depression states that depressed individuals participate in fewer pleasant activities and derive less pleasure and feelings of accomplishment from such activities [6]; in other words, they exhibit reduced behavioral activation (BA). This phenomenon is self-exacerbating: Reduced activation represents a loss of positive feelings that further reduces activation. Neurobiological findings suggest that dysfunction of reward networks (especially in reward valuation, effort valuation, action selection, preference-based decision making, and reward learning) is a central process perpetuating depression [7,8]. For this reason, "reward exposure" aiming to induce BA has been thought to reactivate and retrain reward networks and improve depression [9]. BA therapies are therapeutic approaches based on the relationship between depressive symptomatology and engagement with pleasant activities. They aim to reduce depression symptoms by activating the reward system and have been shown to be as effective as learning-based therapies while being easier to understand for patients and easier to deliver for therapists [6]. An example is the streamlined, evidence-based psychotherapeutic strategy called Engage [7], which aims to systematically address disengagement from participation in pleasurable activities in a structured approach by incorporating reward exposure and addressing barriers in 3 behavioral domains: negativity bias, apathy, and emotional dysregulation. A recent randomized controlled trial showed that Engage is as effective as problem-solving therapy in treating late-life major depression, while having the advantage of being less complex; Engage required 30% less training time compared to problem-solving therapy [10].

To better understand the relationship between BA-based therapies and therapeutic response in depression, robust metrics of the underlying theoretical constructs that minimize reliance on patients' and providers' subjective reports are needed. Text-based therapy sessions provide a unique opportunity to develop such metrics because all language exchanged in these encounters is archived.

Language as an Indicator of Mental State

Language reflects both conscious and subconscious thoughts and feelings [11-13]. Previous work has shown that depressed individuals use language differently than nondepressed individuals in a manner anticipated by cognitive theories of depression. For example, depressed individuals use more first-person singular pronouns (eg, "I," "me," "my") than nondepressed individuals [14,15], indicating increased self-focused attention, a language use consistent with Pyszcynski and Greenberg's [16] integrative model of depression. Depression has also been shown to be associated with a lack of social integration or social disengagement [17-19]. For this reason, Rude et al [14] anticipated a reduction in use of first-person plural pronouns (eg, "we," "us") in depressed college students but had too low a base rate to assess its impact in the sample available for analysis. Stirman and Pennebaker [19] found that suicidal poets used fewer first-person plural pronouns than nonsuicidal poets. Linguistic indicators of positive and negative affect differ in depression and have shown utility in social media–based predictive models of depression [14,20]. These findings are consistent with the emphasis on negative valence in Beck's [21] influential depression theory. Finally, prior work has investigated content word usage by depressed individuals compared to control groups without depression. These include words related to sadness, as well as words related to somatic health concerns (health and biology words, with the biology category combining body, health, sexual, and ingestion words) [22,23]. Given these findings, the question arises whether variations in language use related to the behavioral theory of depression can be detected through natural language processing.

One approach to capturing emotional affect, linguistic style, and topics in written text is to calculate the percentage of words belonging to defined categories, such as positive affect words, pronouns, or words related to certain topics (eg, health or leisure). The Linguistic Inquiry and Word Count (LIWC) software package, a tool developed to study linguistic indicators of mental states, embodies this technique and was used to quantify relevant pronouns and affect words in the aforementioned work. As reviewed by Tausczik and Pennebaker [24], numerous experiments have validated the LIWC categories. However, while LIWC constructs such as "leisure" are related to the notion of activation, they do not provide a comprehensive account of how engagement might manifest in language. For example, categories of relevance to BA, such as the breadth of activities one engages in or the extent to which one derives a sense of accomplishment from setting and achieving activity goals, are not represented in the LIWC standard dictionaries.

Distributional representations of words learned from large amounts of electronic text can help construct comprehensive

XSL•FO
RenderX

sets of terms similar to the curated sets used by LIWC to represent categories. Also known as semantic vectors or word embeddings, these representations are learned from text, with a typical approach involving first initializing random vectors of user-defined dimensionality and then iteratively updating them to make vectors for words appearing in similar contexts similar to one another. With neural embeddings, this is achieved by training a neural network model to predict the words surrounding an observed word and retaining some of the neural network weights after training to serve as word embeddings. Empath [25] is a tool designed to support rapid computer-assisted construction of user-defined term sets using such embeddings to find terms that are similar to an initial set of seed terms. Term sets constructed in this way have a strong correlation with the corresponding LIWC categories, which were constructed in a completely manual process. In essence, Empath's approach uses distributional representations of words to identify similar terms to a set of seed terms based on their distributional statistics across a large text corpus. In this way, a small seed set of terms can be rapidly expanded to provide adequate coverage, with the expanded list provided to manual reviewers for pruning of those terms considered to be inconsistent with the category of interest. Empath's vector representations are derived from a corpus of fiction. Though generally harder to come by, customized in-domain training corpora are known to produce better word representations in clinical domains [26].

For the current work, we developed a metric of BA, using distributional representations derived from a large corpus of naturally occurring language from online therapy chat messages (n=2,527,783) and characterized its relationship to indicators of depression severity. We hypothesized that linguistic markers of activation would be more frequent in milder depression than in severe depression and that longitudinal changes in these markers would reflect the trajectories of patients' depression; patients who improve over time should also show an increase in BA. We further hypothesized that linguistic markers of BA would capture a separate, clinically meaningful dimension of depression symptomatology — namely, engagement in meaningful, rewarding activities — compared to the established linguistic indicators, which capture psychological manifestations of depression (self-focused attention or social integration [function word usage] and emotional tone) and content topics (sadness, health, biology words). Therefore, the BA metric should capture information beyond that reflected by established markers. We tested these hypotheses in the subset of messages from the time period where evaluations of the severity of depression were available for participants at regular intervals (n=1,051,025).

Methods

BA Lexicon

We developed a lexicon of related words collectively representing the construct of activation as used in BA. We constructed a set of 66 unique representative seed terms, informed by the Activation subscale of the Behavioral Activation for Depression Scale (BADS) [27], a validated instrument used to identify subjective engagement levels. The subscale consists of 7 questions, each aiming to capture a unique component of the construct. Seed terms were selected manually for each question in collaboration with GA, a clinician investigator with extensive experience in BA approaches (Table 1).

We expanded the sets of terms for the novel LIWC construct by using methods of distributional semantics, which generate vector representations of words from their distributional statistics in text, such that words occurring in similar contexts will have similar vector representations [28,29]. Specifically, we used the open-source Semantic Vectors software package [30-32] to train 100 dimensional word embeddings using the skipgram-with-negative-sampling algorithm [33] on a set of 2.5 million de-identified messages sent by Talkspace clients (>165 million total words). Embeddings were trained over 10 epochs, using a sliding window radius of 2 and a subsampling frequency threshold of 10^{-5}. Words occurring fewer than 5 times in the corpus were excluded from training. This minimum frequency threshold is employed to restrict model consideration to those terms that occur in a sufficient number of contexts to inform a distributional representation and to constrain the number of vectors to save time (during nearest neighbor search) and disk space. We did not attempt to optimize this parameter but note that it is the default in the canonical implementation of the skipgram-with-negative-sampling algorithm [34]. For each seed term, we then added the 30 most related terms as determined by the cosine similarity between the seed term's vector representation and the vectors for all other terms. We chose to add 30 because this number appeared to achieve high coverage while imposing a manageable workload for manual pruning. Note that stemming is not necessary, as this process will capture all forms of a word appearing in the training text, while preserving their semantic nuances; further, keeping all words appearing in the raw text ensures consistency between our dictionary and the texts to be assessed. For illustrative examples of similar words, see Table 2.

These lists of terms were manually filtered to remove irrelevant or inaccurate terms. Then, we solicited feedback and suggestions from GA. Feedback was incorporated at the seed term and manual filtering steps. The process was iteratively repeated until the lexicon was found to be to be satisfactorily inclusive and specific. Finally, the expanded lists of words, one per seed term, were combined to form 7 partially overlapping subconcepts according to Table 1, as well as one overarching "activation" category. In the overarching activation category, duplicate terms were removed to prevent double counting. We obtained a set of 1059 unique words, which represent the overarching idea of BA. The words originating from each of the 7 items in the activation subscale of the BADS yielded the subconcepts: *satisfaction* (227 words), *breadth* (341 words), *decisions* (205 words), *accomplishment* (154 words), *long-term* planning (240 words), enjoyment of *effort* (342 words), and *structure* (216 words). The complete term sets are available in the accompanying online repository [35]. LIWC was then used to measure the frequency of words belonging to each construct as a metric of patient engagement in BA.

XSL•FO
RenderX

Table 1. Seed terms derived by the authors from the individual questions in the "Activation" subscale of the Behavioral Activation for Depression Scale (BADS).

Item	Brief name assigned by the authors	Derived seed terms[a]
I am content with the amount and types of things I did.	Satisfaction	Accomplish, achieve, satisfaction, satisfied, enjoy, content, contentment, accomplishment, love, proud, inspired, inspiring, enthuse, affirm
I engaged in a wide and diverse array of activities.	Breadth	Activity, active, participate, involved, event, powerlifting, water coloring, exercise, sport, basketball, restaurant, hobby, craft, art, music, instrument, piano
I made good decisions about what type of activities and/or situations I put myself in.	Decisions	Decision, planning, plan, contest, competition, opportunity, chance, spontaneous, whim, spur, attentive, affirm, commit, focus
I was an active person and accomplished the goals I set out to do.	Accomplishment	Goals, accomplish, progress, goal, achieve, effort, content, contentment, accomplishment, proud
I did things even though they were hard because they fit in with my long-term goals for myself.	Long-term	Goals, progress, goal, effort, planning, plan, challenge, attentive, birth, commit, change, invest, life, payoff, benefit
I did something that was hard to do but it was worth it.	Effort	Effort, enjoy, excited, energized, energizing, love, contest, competition, challenge, chance, fun, enthusiastic, inspired, inspiring, enthuse, event, affirm, commit, change, focus, fuel, invest, invigorate
I structured my day's activities.	Structure	Goals, progress, goal, planning, plan, structure, attentive, event, routine, schedule, regular

[a]There are 104 total words in the right column, including duplicates (eg, "goals" appears in accomplishment, long-term, and structure) for a total of 66 unique terms.

XSL•FO
RenderX

Table 2. Examples of seed terms and similar terms with corresponding similarity score, calculated by computing the similarity between word vectors.

Seed terms and similar terms[a]	Similarity score
Proud	
Accomplished	0.729
Accomplishment	0.679
Accomplishments	0.673
Impressed	0.667
Prouder	0.663
Gussied	0.646
Active	
Inactive[b]	0.659
Activity	0.633
Powerlifter	0.607
Motivated	0.605
Mighy	0.600
Intramural	0.592
Decision	
Decisions	0.863
Choice	0.841
Deciding	0.723
Hyphenating	0.697
Choices	0.693
Decide	0.671
Goal	
Goals	0.865
Attainable	0.761
Achievable	0.740
Acheive	0.725
Aim	0.722
Accomplish	0.717
Commit	
Committing	0.828
Committed	0.788
Babydaddy[b]	0.708
Committ	0.708
Commitment	0.706
Sucide[b]	0.682
Effort	
Efforts	0.729
Concerted[b]	0.718
Valiant	0.690
Handsomeness[b]	0.687
Timeand[b]	0.662

XSL•FO
RenderX

Seed terms and similar terms[a]	Similarity score
Independents[b]	0.648
Routine	
Routines	0.874
Schedule	0.708
Nighttime	0.698
Regimen	0.691
Rhythm	0.682
Schefule	0.682

[a]Terms were extracted from our chat message corpus and thus include common typographical errors.

[b]Words that were removed in the filtering process.

LIWC

The LIWC [24,36] software package, developed by Pennebaker and his colleagues over the past 2 decades, was used for linguistic analysis. LIWC derives features from narrative text by counting the number of words in a text that correspond to categories in the LIWC lexicon (or dictionary), with categories defined by lists of words that fall into them. LIWC returns the percentage (or proportion) of words in a text that correspond to each category. For example, consider the following excerpt from an interview with singer, songwriter, and poet Leonard Cohen:

> *When I speak of depression, I speak of a clinical depression that is the background of your entire life, a background of anguish and anxiety, a sense that nothing goes well, that pleasure is unavailable and all your strategies collapse.* [37]

This excerpt is 40 words long, and the words "depression" (n=2), "anguish" (n=1), and "anxiety" (n=1) fall into the LIWC negative emotion category. Therefore, LIWC returns a percentage score of 10% (100*4/40) for this category. Other categories are measured similarly by estimating the frequency with which words they include occur in a unit of text. However, LIWC also includes a set of composite categories that are derived by combining individual categories. As negative and positive affect are both potentially informative, for parsimony, we considered the composite emotional tone variable, which combines the positive and negative emotion categories. A high tone score indicates a predominance of positive over negative emotion words, and a low score indicates the opposite. A score of 50 indicates a balance between positive and negative affect [36].

Additionally, we measured the usage rates of first-person singular pronouns, first-person plural pronouns, and words belonging to the content categories health, biology, and sadness. Finally, we measured linguistic indicators of BA by counting the number of BA lexicon words overall and in each subcategory in every patient's messages.

Data

This work utilized de-identified chat messages sent during routine online therapy, collected for a previously reported study by Hull et al [38]. Clients took part in messaging therapy conducted by a licensed, certified clinical professional via the Talkspace online platform over 12 weeks. The platform provides a paid service open to all, and the service may be covered by some insurers. Therapists and clients converse via written, asynchronous messaging on the platform, and therapists utilize a range of therapeutic strategies. The platform also allows users to send video and audio messages, though these were not used in the current work. Only client messages collected during the course of therapy were used in this study. Participants completed the Patient Health Questionnaire 9-item (PHQ-9) questionnaire at baseline as well as every 3 weeks during therapy. The PHQ-9 is a validated self-report questionnaire commonly used to assess depression severity, scored on a scale of 0-27 [4]. For further details on the platform, data collection process, and study population, see Hull et al [38].

The participants (N=10,718) were young (≤35 years old: 8014/10,142 [576 age values were missing], 79.02%; none younger than 18 years), educated (Bachelor's degree or higher: 6871/9169 [1549 education values were missing], 74.94%), and mostly female (8340/10,571 [147 gender values were missing], 78.90%). Data on race and ethnicity are not systematically collected by the digital platform and are missing for most participants. There were a total of 24,387 PHQ-9 assessments with corresponding messages, with 37.65% of participants (4035/10,718) only completing the baseline assessment, 24.50% (2626/10,718) completing 2 assessments, 18.31% (1962/10,718) completing 3 assessments, 9.69% (1038/10,718) completing 4 assessments, and 9.86% (1057/10,718) completing 5 assessments. The mean baseline PHQ score was 13.36 (SD 4.96) and did not significantly vary with the total number of assessments completed. The mean end PHQ score was 10.80 (SD 5.83) and was significantly lower the more assessments were completed. Patients participated in chat conversations throughout the study period, as well as in the 3 weeks leading up to the baseline assessment in some cases, which were included when available (weeks –3 through –1). Messages were aggregated by concatenating them (ie, combining messages in sequence), creating a single "document" as the unit of analysis. For studies 1 and 2, each PHQ score was used to label the pooled messages from the period on which the questionnaire asks respondents to reflect (the 2 previous weeks). PHQ-9 questionnaires were filled out at the beginning of weeks 0, 3,

XSL•FO
RenderX

6, 9, and 12. For study 3, messages were pooled by week (starting with week –3), and each series of (up to) 15 datapoints has one trajectory label. On average, participants had 7.4 weeks of messages and wrote 770 words per week; patients completed 2.2 assessments on average and wrote 2133 words per completed assessment. At baseline, the number of words written did not vary significantly with depression severity (P=.33). There were 79,096 weeks of messages and 23,950 PHQ assessments with messages. For discussion of the relationship between demographic and engagement factors and treatment outcomes, please see Hull et al [38].

Trajectory Labels

Based on patients' longitudinal PHQ-9 and General Anxiety Disorder-7 (GAD-7) scores, Hull et al [38] clustered patients using latent growth modeling and assigned the following labels to the 6 trajectory groups that emerged: Acute Recovery, Recovery, Depression Improvement, Anxiety Improvement, Chronic, and Elevated Chronic. The middle 2 categories appeared to capture patients who improved in some symptoms but not others. Additionally, improvements in PHQ (or lack thereof) were less clear than in the other groups. Because the individuals in these groups are thus outside the simple definitions of depression "improvement" and "nonimprovement," they were not included in analyses of binary improvement status. A subset of 6760 patients was used for trajectory analysis. Of these patients, which Hull et al [38] identified as strictly "improving" or "nonimproving," 47.2% (3189/6760) improved (classified as Recovery or Acute Recovery), and 52.8% (3571/6760) did not improve (classified as Chronic or Elevated Chronic).

This study focused on ascertaining the utility of linguistic markers to predict depression symptom improvement only; therefore, when using trajectories, we simplified trajectories into a binary "improvement" label, with the 2 recovery classes in the improvement group and the 2 chronic classes in the nonimprovement group.

Statistical Analysis

Study 1: Association of Linguistic Markers With PHQ

To validate the basic premise of LIWC and the BA concept, we investigated the relationship between linguistic markers and PHQ-9 scores using the (up to) 5 measurements of linguistic indicators with corresponding PHQ-9 scores per patient. For this analysis, each pair of PHQ-9 assessment and corresponding message log was treated as a data point. We first determined whether the established LIWC metrics as well as our novel BA metric are statistically significantly different between patients with different depression symptom severity. Severity was defined by the clinical depression symptomatology groups used by the PHQ scoring system: minimal (PHQ ≤4), mild (PHQ=5-9), moderate (PHQ=10-14), moderately severe (PHQ=15-19), and severe (PHQ ≥20). Further, the average difference in each linguistic marker for each unit difference in PHQ score was determined using mixed effects linear regression, treating the patient identity as a random effect.

Study 2: Utility of BA Subconstructs

Each question of the BADS activation subscale aims to capture a distinct dimension of the theoretical construct. To determine the difference between the components and the potential clinical value of the subcomponents compared to pronoun usage, affect measures, and the overall BA concept, we conducted regression analyses on combinations of different variable subsets. For each analysis, we determined the variance explained by each subset of predictors in a mixed effects model with PHQ-9 score as the outcome, treating participant identity as a random effect. Predictors were combinations of (1) subsets of the established LIWC variables (first-person singular pronouns, first-person plural pronouns, emotional tone, or all 3; sadness, health, biology, or all 3) and (2) subsets of the BA variables (the overall construct, each of the 7 subconstructs, all 7 subconstructs, and all 7 subconstructs plus the overall construct). Comparing the amount of variance explained (R^2) between baseline models and models that include additional variables yields insights into the extent to which the added variables provide further information. However, chance associations alone can increase R^2 even if variables provide little usable additional information; therefore, we additionally determined the Akaike information criterion (AIC), which penalizes model fit in response to model complexity. A nonincreased AIC in conjunction with an increased R^2 should therefore signal that added variables contained new information.

Study 3: Association of Linguistic Markers With Patient Trajectories

To determine the association between different linguistic indicators and outcome, mixed effects linear regression analysis was used to compare the rates of change of the variables over time between patient trajectories (whether patients were improving [ie, classified as Recovery or Acute Recovery] or nonimproving [ie, classified as Chronic or Elevated Chronic]).

We compared the average change in each variable per 1-week difference (regression slope). For this analysis, messages were aggregated by week, yielding a time series with up to 15 data points for each patient. Thus, we calculated how PHQ scores and linguistic indicators changed with time in the improving and nonimproving groups, controlling for the within-patient dependency of samples. Specifically, for each of the 2 groups, we fitted a mixed effects linear regression model of the following form:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \gamma_{0i} + \gamma_{1i} X_{ij} + \varepsilon_{ij}$$

Where Y_{ij} is the variable of interest measured for participant i in week j (eg PHQ, activation, or satisfaction), X_{ij} is the week number, β_0 and β_1 are the fixed effect (time) parameters, and γ_{0i} and $\gamma 1i$ are the random effects (participant ID) parameters. Calculations were done using the statsmodels package in Python [39].

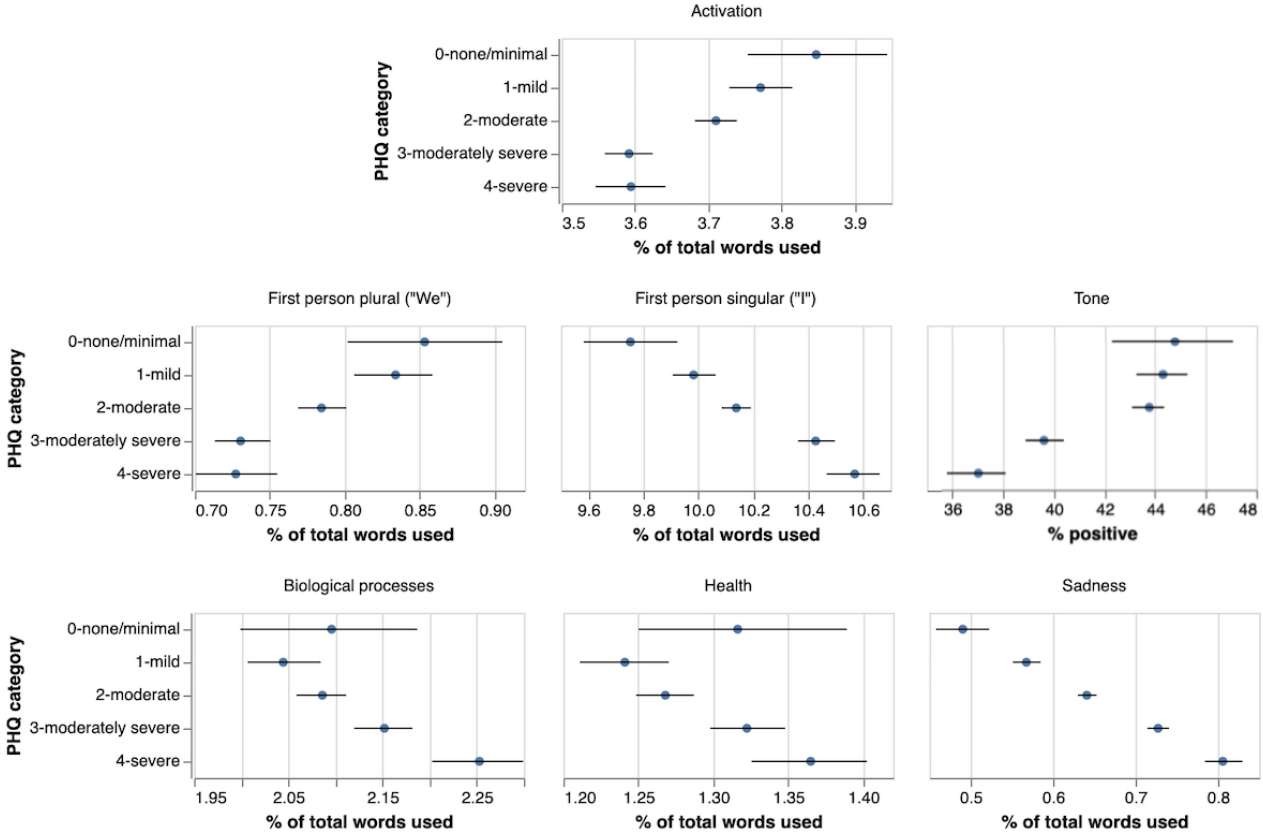
XSL•FO
RenderX

Results

Study 1: Relationship Between Linguistic Indicators and PHQ Scores

Using LIWC to measure the percentage of words belonging to the overall activation construct (including all terms related to any of the subconstructs), the average level of activation across the baseline chat logs was 3.66 (SD 0.89) and varied significantly with the depression symptom severity category (Figure 1), as did the LIWC emotional tone measure and the LIWC pronoun measures (first-person singular and first-person plural). Less depressed individuals used more "we" pronouns and fewer "I" pronouns. All individuals expressed more negative affect than positive (tone <50), with the most depressed individuals exhibiting an emotional tone balance most extremely tipped towards negative affect (lowest scores). The topic-related word categories were also significantly different between severity groups, with sadness having the most pronounced differences between groups. The health and biology categories appear to show increased usage in more depressed individuals but have remarkably large confidence intervals for the least depressed group (none/minimal). Higher overall BA levels were detected for lower depression levels, indicating that patients with more severe depression symptoms discussed activities and associated feelings of enjoyment and reward less than their less-depressed counterparts.

Figure 1. Bootstrapped 95% CIs of the mean of each Linguistic Inquiry and Word Count (LIWC) measure by depression symptom severity category at baseline: minimal (Patient Health Questionnaire [PHQ] ≤4, n=393), mild (PHQ=5-9, n=1865), moderate (PHQ=10-14, n=4109), moderately severe (PHQ=15-19, n=3002), severe (PHQ ≥20, n=1331).

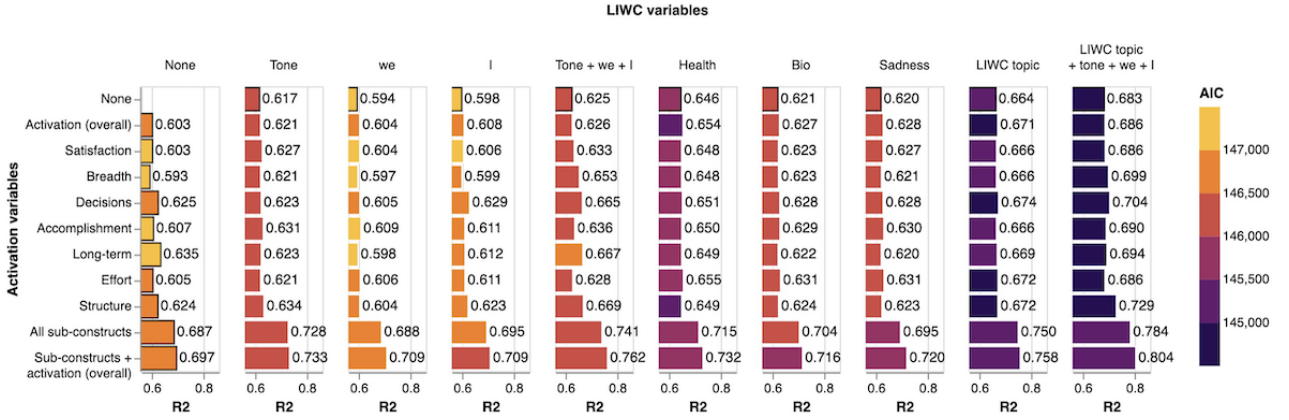


Study 2: Utility of Subconstructs

The variance in the overall PHQ score explained (R^2) by the 109 models fitted to all possible combinations of the LIWC and BA variable sets is shown in Figure 2. The amount of variance explained for each baseline model is shown as bars with strokes, and comparison values are shown without. Darker colors indicate better fit as measured by the AIC. At baseline, emotional tone is more informative than "I" or "we." The health category is the most informative topic category; the 3 topics together explain more variance than the tone and pronoun variables together and also have better fit. All LIWC variables together explain the most variance without detracting from model fit. Of the activation subconstructs, the decision, long-term planning, and daily structure components are most informative; again, all variables together explain the most variance. Including tone added more to satisfaction, breadth, accomplishment, and effort than to decisions, long-term planning, and daily structure. LIWC package constructs alone (tone, I, we; content topics) accounted for 68.3% of the variance, while the combination of our newly created BA subconstructs plus total score alone accounted for 69.7% of the variance. The highest R^2 of 80.4% was achieved by the model that included all variables: emotional tone, function words (I, we), all 3 topic categories, and all 7 subcomponents of activation, along with the overall activation level. Interestingly, including the overall BA concept along with the BA subconstructs appeared to improve both R^2 and fit compared to the subconstructs alone.

XSL•FO
RenderX

Figure 2. Variance explained (R2) by each subset of variables in a mixed effects model with Patient Health Questionnaire (PHQ) score as the outcome. Compare columns to the left column (baseline) for the increase in R2 due to standard Linguistic Inquiry and Word Count (LIWC) variables compared to activation variables alone; compare rows to the top row (baseline) for the increase in R2 due to activation variables compared to standard LIWC variables alone. Darker AIC colors indicate better fit.



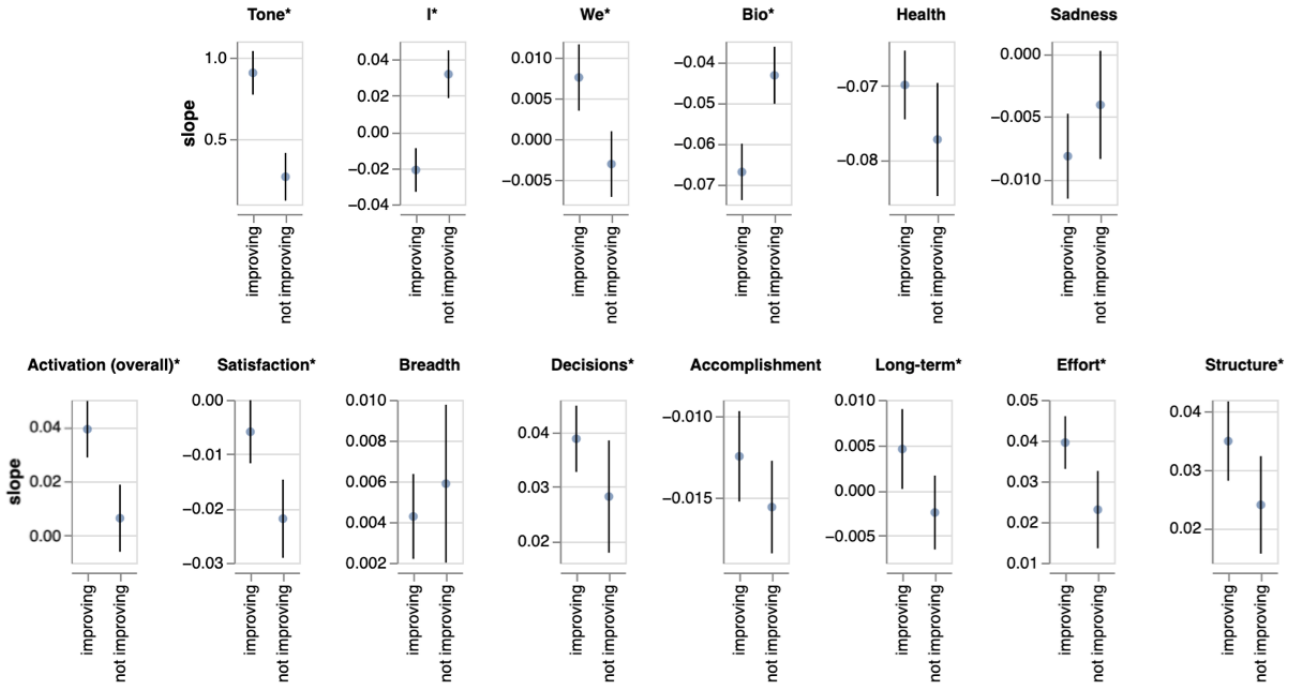
Study 3: Relationship With Patient Trajectories

Figure 3 shows the average change in each linguistic marker per week in the improving and nonimproving groups. Several linguistic indicators showed average amounts of change over time that were significantly different between the 2 groups.

Of the established LIWC markers, emotional tone, first-person singular pronouns, first-person plural pronouns, and biology words were different between groups. Interestingly, biology word usage decreased less in the nonimproving group than in the improved group, while health word usage decreased more in the nonimproving group. Sadness was reduced in both groups over time, with a larger change in the improving group, though the difference between groups was not statistically significant.

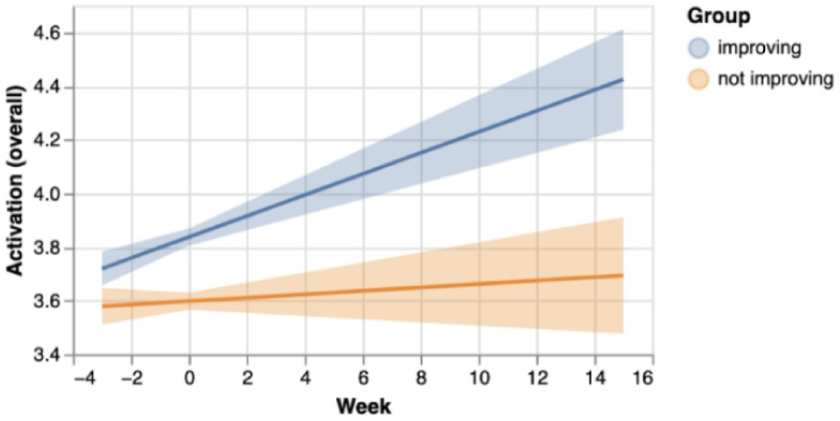
Of the linguistic markers of BA, the markers for satisfaction with activities and rewarding effort had the most pronounced difference between groups, along with the overall activation marker. The fitted fixed effects models are shown in Figure 4. Neither the breadth of activities discussed nor mentions of feelings of accomplishment were different between groups.

Figure 3. Regression coefficients and corresponding 95% CIs of the mixed effects models (ie, the average change in the given variable for each treatment week). *P<.05.



XSL•FO
RenderX

Figure 4. Fixed effects of the fitted linear mixed effects models for the improving and nonimproving groups for activation (overall). Improving: activation = 3.837 + week * 0.039; nonimproving: activation = 3.598 + week * 0.006.



Discussion

In this work, we developed an approach to measuring BA from patients' natural language. We demonstrated that the resulting metric, like scores collected with the validated PHQ-9 depression symptom questionnaire, can discriminate between patients on positive and adverse trajectories. Furthermore, we demonstrated that activation can be measured in terms of its distinct, clinically meaningful subconstructs, each complementary to established linguistic indicators (affect, pronouns, content categories), and has potentially different uses in future research and applications. Additionally, we demonstrated that established linguistic indicators obtained using LIWC are associated with changes in depression symptomatology as measured with the PHQ in a large sample of online therapy clients.

At baseline, linguistic markers captured the depression severity experienced by patients in this study. The established LIWC markers of affect, pronouns, and topic categories were significantly different between PHQ-9 severity groups. Remarkably, sadness and the indicator of self-focus showed nearly identical patterns across groups. Though depression without sadness does occur, sadness is perhaps the most famous marker of depression. The agreement between first-person singular pronouns and sadness confirms the theory of self-focused attention and provides further support for the measurement of self-focus using LIWC. Further, even at baseline, BA was markedly different between depression severity groups. The presence of this association at baseline provides support for the connection between activation and depression symptoms that is the basis for BA therapies.

Notably, not all variables can differentiate all depression severity groups. For example, tone appears to be clearly different between the moderately severe and severe groups at baseline, while activation is not. Conversely, mean activation is different between moderately depressed individuals and those with only minimal symptoms, while tone was not. That not all variables were different between all groups may be related to the polymorphous clinical presentation of depression; patients experience different combinations of symptoms to varying degrees [40]. The heterogeneous presentation of depression suggests that any single metric can only partially capture the clinical severity of depression and underscores the importance of measuring multiple aspects of depressive symptomatology. Combining several complementary markers, each carrying some unique information, can maximize predictive power by capturing different dimensions of depressive psychopathology.

The mechanisms underlying improvement due to BA therapies are not sufficiently well understood. A comparison of different BA approaches found that most result in similar benefits, even though they include slightly different protocols, suggesting that some elements of the various activation approaches may be unnecessary [6]. By investigating the individual components of the activation subscale of the BADS in isolation, we showed that each contributes further information: Considering the individual parts, rather than just the overall idea of activation, explains more variance in regression models without deteriorating model fit (worse model fit might be expected if more variance was explained merely because more variables were included). This finding suggests that the features capture distinct dimensions of activation and that some dimensions may be more important than others. For example, the idea of breadth (corresponding to the BADS item "I engaged in a wide and diverse array of activities.") was least predictive of patient improvement over time.

Manos et al [41] previously showed via factor analysis that a modified version of the breadth item ("I engaged in many different activities") should be retained in a short version of the BADS. However, the current work interpreted the item as capturing activity diversity specifically, counting ideas such as "exercise," "restaurant," and "instrument," similar to the Pleasant Events Schedule (PES) questionnaire. The PES asks respondents for the number of times they participated in activities on a list of 320 options in the past 30 days, including "Playing baseball or softball," "Going to a restaurant," and "Playing a musical instrument" [42]. Manos et al [41] pointed out that PES may not represent the key functional activities of every respondent and therefore may not accurately capture activation. Their rewording of the item removes some of the focus on activity diversity. Our findings support the idea that focusing on activity diversity may not be beneficial for the identification of linguistic markers of BA. Future work might modify the dictionary created here to remove the specific named activities. Further research may also grant insights into

XSL•FO
RenderX

opportunities to refine treatment interventions and protocols. For example, it may be prudent to explore modifying BA therapies to remove any emphasis on activity diversity and focus on the frequency of patients' favorite activities.

Our analysis shows that emotional tone and first-person pronouns (singular and plural) are strongly associated with patient trajectory, consistent with prior work. Both the improving and nonimproving groups shifted to a more positive tone over the treatment period (positive, nonzero slope), which may be an effect of participating in therapy — participants may be focusing on solutions rather than problems in conversations with their therapists. While the sadness category was also strongly associated with patient trajectories, the effect was less pronounced than may have been expected, considering that feelings of sadness are common in depression. Surprisingly, the health topic decreased in usage more in those not on a positive trajectory than those who were, while the "biological processes" topic group decreased more in the improving group. The biological processes topic group contains health, ingestion, sexual, and body words. That patients on the path to recovery discuss these topics less as time passes may indicate that they experience fewer somatic symptoms.

In addition to the overall BA construct, we found several subconstructs strongly associated with patient trajectory. Satisfaction, decisions, long-term planning, effort, and structure all showed significant differences between the improving and nonimproving groups (P<.05). Of all activation subconstructs, breadth and accomplishment were most similar between these groups. However, the addition of breadth to the established LIWC markers resulted in more variance explained without reducing goodness of fit. In fact, breadth was the third most informative single activation construct when included alongside all topic and pronoun variables and tone. In other words, while breadth alone is not explanatory, it is informative in the context of other predictors such as emotional affect. This observation may indicate that discussing activities without positive feelings is not indicative of improvement; activities must also be perceived as rewarding.

Tausczik and Pennebaker [24] first began developing and using LIWC in the 1990s. Counting words belonging to semantic and syntactic categories is simple yet effective, as demonstrated in countless experiments across several fields [14,20,22,24,43,44]. A detailed review of LIWC applications is available elsewhere [24]. Despite this extensive existing work utilizing LIWC, sample sizes have historically been small: The largest sample size in the meta-analysis by Edwards et al [15] of first-person singular pronoun use in depression was 966. Containing chat conversations from over 10,000 individuals over almost 3 years, with over 74 million words, our dataset of naturally occurring language is considerably larger than those used in previous experiments validating the relationship between LIWC variables and depression. Our results provide further validation of the LIWC tone- and pronoun-related variables in the context of scores from a validated, standardized instrument for measurement of symptomatology. Changes in established LIWC variables are consistent with case-control differences demonstrated in prior research [14,19,20], with tone and first-person plural pronouns increasing as symptoms decrease and first-person singular pronoun usage decreasing with improvement in symptoms. Therefore, we showed that these metrics indeed reflect depression symptom status in this dataset, providing further strong support for their relationship with depression. However, in the context of the current data, BA variables explained more of the variance in the PHQ data than pre-existing LIWC constructs.

Limitations

Reliance on the PHQ-9 is a limitation because this instrument focuses exclusively on symptoms of depression. Future studies may use instruments assessing well-being and social adjustment.

The trajectories used to categorize patients as improving or otherwise were assigned via unsupervised learning and do not directly correspond to the total change in PHQ over the course of treatment. Rather, they account for the entire series of depression and anxiety scores over the treatment period. A participant may have an absolute decrease in PHQ score (eg, 16, 18, 15, corresponding to a total change in PHQ of 1 point between the beginning and end of treatment); however, the patient may not be experiencing clinically meaningful progress. Thus, depending on the entirety of the PHQ-9 as well as GAD-7 scores over the treatment period, the model may not assign an "improvement" category for such a patient. Many patients only had 1 or 2 scores, and some participants with identical sets of scores were assigned to different groups. However, mixed effects linear regression confirmed that depression symptoms as measured by PHQ-9 scores improve significantly more for patients in the improving group than for patients in the nonimproving group. Thus, we believe that the categories are sufficiently accurate for our purposes. While we excluded the "gray area" trajectories of Depression Improvement and Anxiety Improvement, it is worth noting that analyses were repeated with these included, and results, though less clear, were not different and our conclusions held; this is the expected result of introducing additional noise. Another potential limitation of the trajectories is that even the "nonimproving" group showed a slope in PHQ that was significantly different from 0; we believe that this is an effect of unsupervised learning (ie, the trajectories clustered into groups labeled as "chronic" are distinct from the other clusters, but not necessarily entirely without improvement) coupled with the fact that therapy and just the simple passage of time (regression to the mean) is at least somewhat effective for most people. A truly "nonimproving" control group is therefore difficult to carve out in this dataset, possibly because such people represent a minority of participants and were thus not assigned a separate cluster by the unsupervised approach. Future work on this dataset may consider using the more granular patient trajectories if the degree of improvement or exacerbation is of interest. For example, a patient may be responding well ("remission"), but there may be room for improvement ("acute remission"); alerting providers to this situation could result in additional efficiency of care.

The lexicon of words collectively representing activation developed in this work was based solely on the messages in this dataset. This approach ensured that the concept represents language usage in our specific study population. While domain-specific texts generally work better for distributional

XSL•FO
RenderX

semantics than more general corpora, they may also result in artifacts with limited generalizability. Our study population was predominantly young (8014/10,718, or 74.77%, were 35 years old or younger) and, considering that they used a paid online therapy service, presumably financially stable; in addition, 2679 (2679/10,718, 25.00%) were residents of California or New York, and 8340 (8340/10,718, 77.81%) were female. This striking gender imbalance is consistent with previously reported gender imbalances in both online and face-to-face therapy. For example, Chester and Glass [45] reported that 70% of online therapy clients are female, and a recent comparable online mental health service in Australia [46] reportedly had 72% female clients. Sagar-Ouriaghli et al [47] reported that women are 1.6 times more likely to receive any form of mental health treatment than men. Nevertheless, the makeup of our study population must be considered in future applications of our results. Geographically and demographically diverse groups may significantly differ in their word choice and usage; as a result, the lexicon may not be generalizable to other groups.

Word count approaches such as the one employed by LIWC have limitations. While most modern natural language processing methods account for negations, counting words does not. LIWC has a separate "negation" category, but it only considers single words and thus does not assign negation statuses to individual concepts. For example, describing having planned the day's activities and *not* having planned the day's activities would both count towards our *structure* concept in the same way. Because both statements reflect that the patient engaged with the idea of planning their day's activities, this is arguably a minor limitation. Still, our results indicate that context is more important for some concepts than others; for example, breadth is only informative when considered alongside other markers such as emotional affect to contextualize it. In our future work, we plan on utilizing approaches that can account for negations.

A potential limitation is that the therapy sessions in this work did not specifically use BA therapy. However, BA is a common pathway of many therapies, which through a variety of interventions, increase exposure of depressed patients to meaningful, rewarding experiences.

An important factor to consider when proposing such automated analyses is the degree to which passive monitoring of therapist-client communication for the purpose of measuring BA may be construed as invasive or intrusive. While we did not directly engage with patients in the current work, we note that our recent work in the suicide prevention domain provides some indication that passive monitoring of this sort may be acceptable to patients when conducted by a trusted party, with 68% of survey participants indicating that the automated analysis of personalized web search data for suicide prevention would be acceptable provided this triggered minimally invasive interventions (such as connection to a support network or therapist) only [48].

Another limitation of the study is the absence of qualitative review that could examine the accuracy of our ratings.

Future Work

An additional implication of the potentially causative mechanism of activation is that it should occur before symptom improvement. While more direct measures of sentiment and mindset reflect an individual's current thoughts and feelings, "activity" topic analysis may reveal long-term dimensions of patient trajectory. Discussing activities, plans for activities, or even avoidance of activities shows that a patient engages with the ideas surrounding BA and may indicate that a patient is moving towards or already part of a positive feedback loop of self-perpetuating improvements. Consequently, one may expect metrics of activation to predict longer-term changes more accurately than word use analysis, which may be confounded by the mood of the moment, and thus capture a separate and clinically meaningful dimension of symptomatology and treatment success.

Our results show that the BA subconstructs of *decisions* (independent decision making), *long-term* (planning for the long term and acting accordingly), and *structure* (structuring daily activities) are parallel in terms of the amount of variance they explain. Let us call these *activities*. Similarly, *satisfaction*, *accomplishment* (a sense of having accomplished something), and *effort* (enjoyment due to effort exerted) are similar to each other. Let us call these *reactions*. Activities alone were more explanatory of improvement than reactions alone; in other words, activities are explanatory even without considering emotional tone. Adding tone increased the amount of variance explained for by reactions to the levels for activities alone. In other words, reactions are more informative when considered alongside tone than without, whereas adding tone makes comparatively little difference for activities. A possible explanation is that merely discussing feelings of satisfaction, or lack thereof, may not correspond to improvement. However, engaging in long-term planning, scheduling activities, and structuring routines does correspond to symptom improvement. Mentions of being satisfied may only be triggered positively once the positive feedback loop is set in motion. Conversely, this pattern is in agreement with the helplessness theory [16,49], which states that experiencing negative consequences to reward-seeking behavior perpetuates the avoidance phenomenon and therefore, continuing depression symptoms, by supporting a negative outlook on the future. Therefore, it requires additional information about the patient's tone to be informative of current symptom severity. Therefore, an opportunity for future work is to investigate the temporal relationship between symptom improvement and the constructs in the activities group compared to those in the reactions group. One may be more useful than the other to predict long-term changes. Ascertaining the timeline of changes in these different activation metrics relative to patient trajectories is a prerequisite to the translation of linguistic indicators into clinical insights.

Recently, a broad range of sophisticated natural language processing techniques has gained traction. Our work here demonstrated that theme and affect analysis conducted via established, straightforward methods such as word counting both reflect and potentially predict depression status. It is plausible that more advanced methods may surpass those used here. Consequently, future work is set to include a deep neural

XSL•FO
RenderX

network model trained on GoEmotions [50], a large corpus of social media posts annotated for the extent to which they express a set of emotion categories [51].

Having validated the linguistic indicators of depression and activation presented here, the question of how to incorporate them effectively into care processes remains. Predictive analytics solutions must be operationalized effectively to improve health outcomes. Accordingly, future work should focus on testing and iteratively refining these measures as part of care delivery. For example, metrics of depression symptom severity and BA may be automatically extracted from patient messages on virtual chat therapy platforms and used to guide the therapist in recognizing whether a patient responds well to therapy or if a change in direction is appropriate.

Conclusion

This work makes several key contributions. First, we devised a computational method to automatically assess theoretical constructs of BA from patient language. Second, building on prior work demonstrating that activation has a close relationship with depression scores, we demonstrated this new metric reflects depression symptom severity. Third, we validated established linguistic markers of depression in a large corpus of naturally occurring language collected as part of psychotherapy sessions, presenting differences between participants with low and high PHQ-9 scores. Fourth, we showed that both the well-established LIWC measures as well as the novel BA measures have utility in predicting longitudinal patient trajectories. Finally, we demonstrated that our metrics of the individual subconstructs of BA capture distinct dimensions of the underlying mechanisms and may lend themselves to unique clinical insights. This work therefore enables further work in automated diagnosis and assessment of depression, as well as refinement of BA psychotherapeutic strategies.

Acknowledgments

This work was supported by the National Library of Medicine (grant number 67-3780; HAB) and the National Institutes of Mental Health (grant numbers R01 MH102252, P50 MH113838; GSA and grant number P50 MH115837; MDP, PAA).

Conflicts of Interest

TDH is an employee of the platform that provided the data. GSA served on the Advisory Boards for Eisai and Janssen Pharmaceuticals. He also served on the Speaker Bureaus of Allergan, Otsuka, and Takeda-Lundbeck. The other authors declare no conflicts of interest.

References

1. Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm [accessed 2021-07-04]
2. Scott K, Lewis CC. Using Measurement-Based Care to Enhance Any Treatment. Cogn Behav Pract 2015 Feb;22(1):49-59 [FREE Full text] [doi: 10.1016/j.cbpra.2014.01.010] [Medline: 27330267]
3. Lambert MJ. Is It Time for Clinicians to Routinely Track Patient Outcome? A Meta-Analysis. Clinical Psychology: Science and Practice 2003 Sep 01;10(3):288-301. [doi: 10.1093/clipsy/bpg025]
4. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
5. Choi BCK, Pak AWP. A catalog of biases in questionnaires. Prev Chronic Dis 2005 Jan;2(1):A13 [FREE Full text] [Medline: 15670466]
6. Mazzucchelli T, Kane R, Rees C. Behavioral activation treatments for depression in adults: A meta-analysis and review. Clinical Psychology: Science and Practice 2009 Dec;16(4):383-411. [doi: 10.1111/j.1468-2850.2009.01178.x]
7. Alexopoulos GS, Arean P. A model for streamlining psychotherapy in the RDoC era: the example of 'Engage'. Mol Psychiatry 2014 Jan 26;19(1):14-19 [FREE Full text] [doi: 10.1038/mp.2013.150] [Medline: 24280983]
8. Dillon DG, Rosso IM, Pechtel P, Killgore WDS, Rauch SL, Pizzagalli DA. Peril and pleasure: an rdoc-inspired examination of threat responses and reward processing in anxiety and depression. Depress Anxiety 2014 Mar 21;31(3):233-249 [FREE Full text] [doi: 10.1002/da.22202] [Medline: 24151118]
9. Lewinsohn PM, Graf M. Pleasant activities and depression. J Consult Clin Psychol 1973 Oct;41(2):261-268. [doi: 10.1037/h0035142] [Medline: 4147832]
10. Alexopoulos GS, Raue PJ, Banerjee S, Marino P, Renn BN, Solomonov N, et al. Comparing the streamlined psychotherapy "Engage" with problem-solving therapy in late-life major depression. A randomized clinical trial. Mol Psychiatry 2020 Jul 01:1. [doi: 10.1038/s41380-020-0832-3] [Medline: 32612251]
11. Pennebaker JW. The secret life of pronouns: What our words say about us. New York, NY: Bloomsbury Press; 2011.
12. Campbell RS, Pennebaker JW. The secret life of pronouns: flexibility in writing style and physical health. Psychol Sci 2003 Jan 06;14(1):60-65. [doi: 10.1111/1467-9280.01419] [Medline: 12564755]

XSL•FO
RenderX

13. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003 Feb;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
14. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cognition & Emotion 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
15. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. Journal of Research in Personality 2017 Jun;68:63-68 [FREE Full text] [doi: 10.1016/j.jrp.2017.02.005]
16. Pyszczynski T, Greenberg J. Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. Psychological Bulletin 1987;102(1):122-138. [doi: 10.1037/0033-2909.102.1.122]
17. Schwarzbach M, Luppa M, Sikorski C, Fuchs A, Maier W, van den Bussche H, et al. The relationship between social integration and depression in non-demented primary care patients aged 75 years and older. J Affect Disord 2013 Feb 20;145(2):172-178. [doi: 10.1016/j.jad.2012.07.025] [Medline: 22871528]
18. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Social Science & Medicine 2000 Sep;51(6):843-857. [doi: 10.1016/S0277-9536(00)00065-4]
19. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]
20. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. 2013 Presented at: Seventh International AAAI Conference on Weblogs and Social Media; July 8-11, 2013; Cambridge, MA.
21. Beck A. Depression. Clinical, experimental and theoretical aspects. J R Coll Gen Pract 1967;18(87):249.
22. Sonnenschein AR, Hofmann SG, Ziegelmayer T, Lutz W. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn Behav Ther 2018 Jul 18;47(4):315-327. [doi: 10.1080/16506073.2017.1419505] [Medline: 29345528]
23. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
24. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
25. Fast E, Chen B, Bernstein MS. Empath: Understanding Topic Signals in Large-Scale Text. 2016 Presented at: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA.
26. Roberts K. Assessing the Corpus Size vs. Similarity Trade-off for Word Embeddings in Clinical NLP. 2017 Presented at: Clinical Natural Language Processing Workshop; December 11, 2016; Osaka, Japan.
27. Kanter JW, Mulick PS, Busch AM, Berlin KS, Martell CR. The Behavioral Activation for Depression Scale (BADS): Psychometric Properties and Factor Structure. J Psychopathol Behav Assess 2006 Oct 25;29(3):191-202. [doi: 10.1007/s10862-006-9038-5]
28. Cohen T, Widdows D. Empirical distributional semantics: methods and biomedical applications. J Biomed Inform 2009 Apr;42(2):390-405 [FREE Full text] [doi: 10.1016/j.jbi.2009.02.002] [Medline: 19232399]
29. Turney PD, Pantel P. From Frequency to Meaning: Vector Space Models of Semantics. JAIR 2010 Feb 27;37:141-188. [doi: 10.1613/jair.2934]
30. Widdows D, Ferraro K. Semantic Vectors: A scalable open source package and online technology management application. 2008 Presented at: Sixth International Conference on Language Resources and Evaluation (LREC'08); May 28-30, 2008; Marrakech, Morocco. [doi: 10.1057/9781137294678.0479]
31. Widdows D, Cohen T. The Semantic Vectors Package: New Algorithms and Public Tools for Distributional Semantics. 2010 Presented at: IEEE Fourth International Conference on Semantic Computing; September 22-24, 2010; Pittsburgh, PA. [doi: 10.1109/icsc.2010.94]
32. Cohen T, Widdows D. semanticvectors/semanticvectors. GitHub. URL: https://github.com/semanticvectors/semanticvectors [accessed 2021-07-04]
33. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. arxiv. Preprint published on October 16, 2013 [FREE Full text]
34. Mikolov T. tmikolov / word2vec. GitHub. 2013. URL: https://github.com/tmikolov/word2vec [accessed 2021-07-04]
35. Burkhardt, HA. hannahburkhardt / linguistic-markers-of-ba. GitHub. URL: https://github.com/hannahburkhardt/linguistic-markers-of-ba [accessed 2021-07-04]
36. Pennebaker JW, Booth RJ, Francis ME. Linguistic Inquiry and Word Count: LIWC2007. URL: http://www.gruberpeplab.com/teaching/psych231_fall2013/documents/231_Pennebaker2007.pdf [accessed 2021-07-04]
37. Lynskey D. Leonard Cohen: 'All I've got to put in a song is my own experience'. The Guardian. 2012 Jan 19. URL: https://www.theguardian.com/music/2012/jan/19/leonard-cohen [accessed 2021-07-04]
38. Hull TD, Malgaroli M, Connolly PS, Feuerstein S, Simon NM. Two-way messaging therapy for depression and anxiety: longitudinal response trajectories. BMC Psychiatry 2020 Jun 12;20(1):297 [FREE Full text] [doi: 10.1186/s12888-020-02721-x] [Medline: 32532225]
39. Seabold S, Perktold J. statsmodelsconometric and statistical modeling with python. Inth Python in Science Conference 2010:E. [doi: 10.25080/majora-92bf1922-011]

XSL•FO
RenderX

40. Lewinsohn PM, Sullivan JM, Grosscup SJ. Changing reinforcing events: An approach to the treatment of depression. Psychotherapy: Theory, Research & Practice 1980;17(3):322-334. [doi: 10.1037/h0085929]
41. Manos RC, Kanter JW, Luo W. The behavioral activation for depression scale-short form: development and validation. Behav Ther 2011 Dec;42(4):726-739. [doi: 10.1016/j.beth.2011.04.004] [Medline: 22036000]
42. MacPhillamy DJ, Lewinsohn PM. The pleasant events schedule: Studies on reliability, validity, and scale intercorrelation. Journal of Consulting and Clinical Psychology 1982 Jun;50(3):363-380. [doi: 10.1037/0022-006X.50.3.363]
43. De Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France. [doi: 10.1145/2464464.2464480]
44. Molendijk ML, Bamelis L, van Emmerik AA, Arntz A, Haringsma R, Spinhoven P. Word use of outpatients with a personality disorder and concurrent or previous major depressive disorder. Behav Res Ther 2010 Jan;48(1):44-51. [doi: 10.1016/j.brat.2009.09.007] [Medline: 19819423]
45. Chester A, Glass CA. Online counselling: a descriptive analysis of therapy services on the Internet. British Journal of Guidance & Counselling 2006 May;34(2):145-160. [doi: 10.1080/03069880600583170]
46. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. The first 30 months of the MindSpot Clinic: Evaluation of a national e-mental health service against project objectives. Aust N Z J Psychiatry 2017 Dec;51(12):1227-1239. [doi: 10.1177/0004867416671598] [Medline: 27733709]
47. Sagar-Ouriaghli I, Godfrey E, Bridge L, Meade L, Brown JSL. Improving Mental Health Service Utilization Among Men: A Systematic Review and Synthesis of Behavior Change Techniques Within Interventions Targeting Help-Seeking. Am J Mens Health 2019 Oct 27;13(3):1557988319857009-1557988319851286 [FREE Full text] [doi: 10.1177/1557988319857009] [Medline: 31184251]
48. Areán PA, Pratap A, Hsin H, Huppert T, Hendricks K, Heagerty P, et al. Perceived Utility and Characterization of Personal Google Search Histories to Detect Data Patterns Proximal to a Suicide Attempt in Individuals Who Previously Attempted Suicide: Pilot Cohort Study. J Med Internet Res 2021 May 06;23(5):e27918 [FREE Full text] [doi: 10.2196/27918] [Medline: 33955838]
49. Abramson LY, Seligman ME, Teasdale JD. Learned helplessness in humans: Critique and reformulation. Journal of Abnormal Psychology 1978;87(1):49-74. [doi: 10.1037/0021-843x.87.1.49]
50. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S. GoEmotions: A Dataset of Fine-Grained Emotions. Cornell University. 2020. URL: http://arxiv.org/abs/2005.00547 [accessed 2021-07-04]
51. Cowen AS, Keltner D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proc Natl Acad Sci U S A 2017 Sep 19;114(38):E7900-E7909 [FREE Full text] [doi: 10.1073/pnas.1702247114] [Medline: 28874542]

Abbreviations

AIC: Akaike information criterion
BA: behavioral activation
BADS: Behavioral Activation for Depression Scale
GAD-7: General Anxiety Disorder-7
LIWC: Linguistic Inquiry and Word Count
PES: Pleasant Events Schedule
PHQ-9: Patient Health Questionnaire 9-item

Edited by G Eysenbach; submitted 25.02.21; peer-reviewed by C Entwistle, M Rawsthorne, DA Rohani, A Ferrario; comments to author 19.03.21; revised version received 03.05.21; accepted 31.05.21; published 14.07.21.

Please cite as:
Burkhardt HA, Alexopoulos GS, Pullmann MD, Hull TD, Areán PA, Cohen T
Behavioral Activation and Depression Symptomatology: Longitudinal Assessment of Linguistic Indicators in Text-Based Therapy Sessions
J Med Internet Res 2021;23(7):e28244
URL: https://www.jmir.org/2021/7/e28244
doi:10.2196/28244
PMID:34259637

©Hannah A Burkhardt, George S Alexopoulos, Michael D Pullmann, Thomas D Hull, Patricia A Areán, Trevor Cohen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

XSL•FO
RenderX

Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Intervention for Primary Care Practitioners to Support Antidepressant Discontinuation (Advisor for Health Professionals): Development Study

Hannah Bowers[1], PhD; Tony Kendrick[1], FRCGP; Nadja van Ginneken[2], MRCGP; Marta Glowacka[3], PhD; Samantha Williams[1], MRES; Geraldine M Leydon[1], PhD; Carl May[4], FRCGP; Christopher Dowrick[2], FRCGP; Joanna Moncrieff[5], MD; Chris F Johnson[6], MRES; Michael Moore[1], FRCGP; Rebecca Laine[1], BMBS; Adam W A Geraghty[1], PhD

[1]Primary Care, Population Sciences and Medical Education, University of Southampton, Southampton, United Kingdom

[2]Department of Primary Care and Mental Health, University of Liverpool, Liverpool, United Kingdom

[3]Faculty of Health and Social Sciences, Bournemouth University, Bournemouth, United Kingdom

[4]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom

[5]Division of Psychiatry, Faculty of Brain Sciences, University College London, London, United Kingdom

[6]Pharmacy & Prescribing Support Unit, Pharmacy Services, NHS Greater Glasgow & Clyde, Glasgow, United Kingdom

Corresponding Author:
Hannah Bowers, PhD
Primary Care, Population Sciences and Medical Education
University of Southampton
Aldermoor
Southampton, SO15 5HT
United Kingdom
Phone: 44 2380591743
Email: h.m.bowers@soton.ac.uk

Abstract

Background: The number of people receiving antidepressants has increased in the past 3 decades, mainly because of people staying on them longer. However, in many cases long-term treatment is not evidence based and risks increasing side effects. Additionally, prompting general practitioners (GPs) to review medication does not improve the rate of appropriate discontinuation. Therefore, GPs and other health professionals may need help to support patients discontinuing antidepressants in primary care.

Objective: This study aims to develop a digital intervention to support practitioners in helping patients discontinue inappropriate long-term antidepressants (as part of a wider intervention package including a patient digital intervention and patient telephone support).

Methods: A prototype digital intervention called Advisor for Health Professionals (ADvisor HP) was planned and developed using theory, evidence, and a person-based approach. The following elements informed development: a literature review and qualitative synthesis, an in-depth qualitative study, the development of guiding principles for design elements, and theoretical behavioral analyses. The intervention was then optimized through think-aloud qualitative interviews with health professionals while they were using the prototype intervention.

Results: Think-aloud qualitative interviews with 19 health professionals suggested that the digital intervention contained useful information and was readily accessible to practitioners. The development work highlighted a need for further guidance on drug tapering schedules for practitioners and clarity about who is responsible for broaching the subject of discontinuation. Practitioners highlighted the need to have information in easily and quickly accessible formats because of time constraints in day-to-day practice. Some GPs felt that some information was already known to them but understood why this was included. Practitioners differed in their ideas about how they would use ADvisor HP in practice, with some preferring to read the resource in its entirety and others wanting to *dip in and out* as needed. Changes were made to the wording and structure of the intervention in response to the feedback provided.

Conclusions: ADvisor HP is a digital intervention that has been developed using theory, evidence, and a person-based approach. The optimization work suggests that practitioners may find this tool to be useful in supporting the reduction of long-term

XSL•FO
RenderX

antidepressant use. Further quantitative and qualitative evaluation through a randomized controlled trial is needed to examine the feasibility, effectiveness, and cost-effectiveness of the intervention.

(J Med Internet Res 2021;23(7):e25537) doi:10.2196/25537

KEYWORDS

antidepressant discontinuation; intervention development; depression; primary care; digital intervention

Introduction

Background

Antidepressant prescribing has increased in England every year for more than 30 years [1], and Public Health England estimated that 4.5 million people (around 1 in 10 of the adult population) were taking an antidepressant in March 2018 [2]. This rise is mainly because of an increase in long-term use [3,4]—approximately 40% of antidepressant users have been on them for 2 years or more and 24% for at least 3 years [2].

Although long-term use of antidepressants is considered necessary in some cases to prevent relapse, surveys suggest that approximately one-third to a half of patients on long-term antidepressants have no evidence-based indication for continuing long-term use [5-7]. Attempts to discontinue antidepressants in these patients may improve their quality of life by avoiding adverse effects that tend to increase with longer-term use [8] and may also help to reduce the costs of prescriptions and consultations to the National Health Service [9].

Previous research has suggested that patients may continue to take antidepressants long term because of a lack of general practitioner's (GP) review [10]; however, studies on prompting GPs to review patients on long-term antidepressants have resulted in only 6% to 8% of patients discontinuing, which did not significantly differ from usual care [11,12]. The barriers to stopping antidepressants are multifaceted, and thus, a complex approach to support withdrawal is needed, involving patients, primary care practitioners, and additional psychological support.

Qualitative studies on the long-term use of antidepressants have highlighted a number of barriers to discontinuation, such as the fear of relapse and beliefs about the necessity of long-term use [13]. Importantly, these studies have also highlighted the critical role of the GP-patient relationship and how this can help or hinder discontinuation. However, there is comparatively less research exploring the perspectives and experiences of health professionals (ie, GPs, nurses, pharmacists, and psychological practitioners) regarding antidepressant discontinuation. In a recent qualitative study, Bowers et al [14] highlighted key factors that served as barriers to GPs supporting patients to discontinue long-term antidepressants. These included uncertainty about the responsibility of initiating discussions about discontinuation, with GPs often suggesting it was the patients' responsibility to start these conversations. Fear about destabilizing patients who were well was also a factor, alongside a perceived lack of information and support for safe discontinuation. Importantly, it may be possible to intervene to address such issues, thereby supporting successful discontinuation, where appropriate.

Objectives

The REDUCE (Reviewing Long Term Antidepressant Use by Careful Monitoring in Everyday Practice) program aims to develop and trial safe, feasible, and effective ways to support antidepressant discontinuation. In this paper, we describe the methods involved in the development of a digital intervention aimed at changing the behavior of health professionals involved in antidepressant discontinuation as part of the REDUCE program. We then discuss how this method was implemented and the findings of the development work.

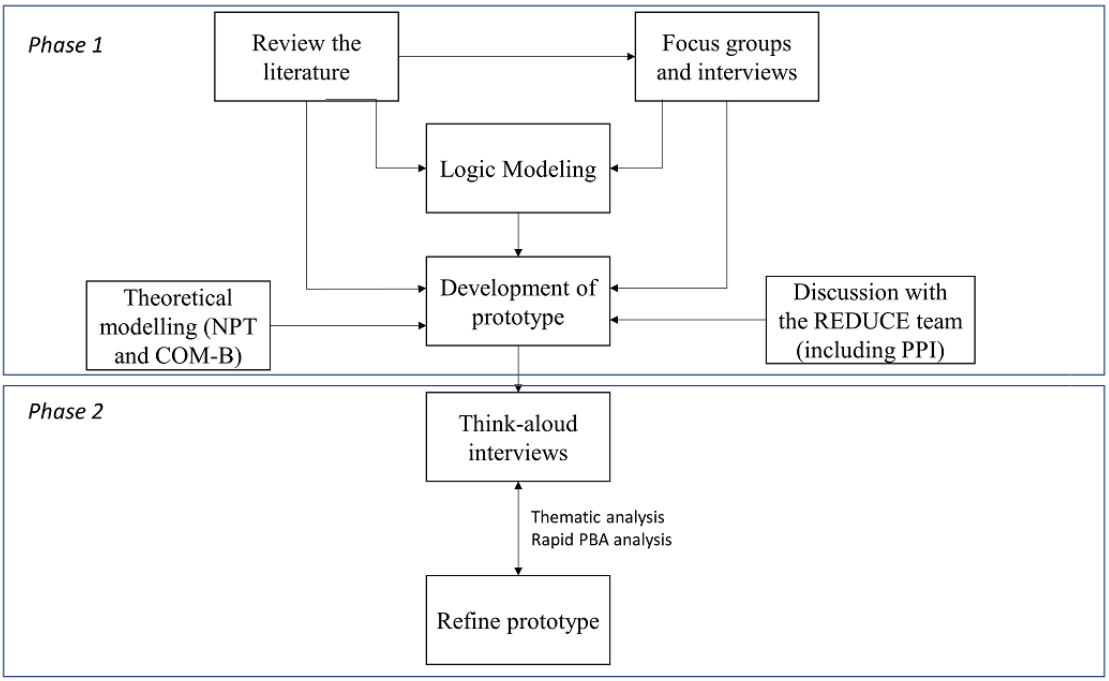
Methods

The Person-Based Approach

The mixed methods approach used to guide the development of the intervention for health care professionals takes commonly applied theory and evidence-based methods and integrates the person-based approach (PBA) [15]. The PBA is a systematic approach for integrating in-depth qualitative research across multiple stages of the development process. It involves the integration of open-ended qualitative interviews, qualitative meta-syntheses, and think-aloud qualitative optimization interviews with relevant theory and quantitative evidence base (Figure 1). Here, we will describe the key elements of this approach, including how we drew from theory (through behavioral analysis and logic modeling) and systematic reviews and conducted primary qualitative research to develop our intervention.

XSL•FO
RenderX

Figure 1. Intervention development method. COM-B: Capability, Opportunity, Motivation-Behavior; NPT: Normalization Process Theory; PBA: person-based approach; PPI: patient and public involvement; REDUCE: Reviewing Long Term Antidepressant Use by Careful Monitoring in Everyday Practice.



Phase 1: Intervention Planning

Step 1: Systematic Reviewing

A qualitative thematic synthesis of 22 studies was conducted, as described in detail elsewhere [13]. In this thematic synthesis, our team identified barriers and facilitators to antidepressant discontinuation from both patient and health professional perspectives. Barriers and facilitators from health professionals' perspectives were then used to inform the guiding principles (Textbox 1) and content for the intervention.

Textbox 1. Guiding principles for the intervention package: design objectives and key intervention features relevant to each design objective.

To persuade and inform health professionals about the reasons and benefits of discontinuing antidepressants

- Ensure that the material covers patients' motivations for stopping
- Include evidence that antidepressants can be withdrawn without harm

To provide rapidly accessible information designed to increase self-efficacy for effectively managing discontinuation in patients

- Simple, unobtrusive interface, using bulleted text as far as possible
- Offer a range of strategies for identifying and managing withdrawal effects
- Use of clear, visually mapped tapering schedules
- Acknowledge flexibility in working with patients

To enable health professionals to support a wide range of patients and be used in a manner that best suits the health professionals

- Provide content that will be relevant for patients in a wide variety of contexts
- Enable use in multiple ways (eg, as a linear educational system or as a resource that can be dipped in and out of), as and when the health professional deems necessary

Step 2: Primary Qualitative Research

We drew from our previously conducted qualitative study with health professionals regarding antidepressant discontinuation (refer to the study by Bowers et al [14] for full paper). This qualitative work comprised 4 focus groups and 3 individual interviews with 38 health care professionals. The focus groups and interviews explored views on long-term antidepressant use, negotiating the decision to discontinue with patients, their role in supporting patients, and important elements of the proposed interventions (including content for practitioners). Professionals included 21 GPs, 4 GP assistants, 7 nurse practitioners (NPs), and 6 mental health workers or psychological therapists. We used themes identified in the interviews and focus groups to inform the content and mode of presentation or delivery of the intervention.

XSL•FO
RenderX

Step 3: Development of Guiding Principles

As part of the PBA, guiding principles—a set of key assumptions followed throughout the process of the intervention development—were formulated. They included broad design objectives that inform how the core intervention strategies are applied and implemented with the aim to increase engagement [16]. The qualitive work (ie, focus groups, interviews, and qualitative metasynthesis) informed the formulation of the guiding principles for the health professional intervention.

Step 4: Behavioral Analysis and Logic Modeling

We drew on implementation and behavioral theory in addition to our qualitative work and discussions with experts (including patient representatives, GPs, psychiatrists, psychologists, sociologists, and health service researchers) to identify key components of the intervention. We conducted a behavioral diagnosis using the Behavior Change Wheel (BCW) and the COM-B Model (capability, opportunity, motivation and behavior) of behavior change [17,18] to identify key behaviors that needed to be addressed or targeted in the intervention and their determinants or context. Our proposed ideas about intervention content were then mapped theoretically using the BCW and Normalization Process Theory (NPT) [19]. NPT is a framework for understanding the key mechanisms that shape implementation processes. It focuses on the work that participants do to incorporate components of complex interventions into their everyday lives [19]. In addition, we used logic modeling to draft how the intervention may work to effectively support appropriate discontinuation. Logic models [20] represent theories of change in a diagrammatic form. They are a useful tool to highlight assumptions regarding potentially important mechanisms and are adapted as data become available.

Step 5: Drafting Intervention Content

The intervention's content was developed using findings from the qualitative metasynthesis of the literature, primary qualitative work, behavioral analysis, and logic modeling. Written intervention content was drafted by a member of the content development team (HB) using what was learned from the qualitative work and theoretical modeling. These drafts were then discussed by the content development team (HB, AG, and MG) and then the wider team (which included patients, GPs, sociologists, psychologists, and psychiatrists). Discussions led to iterations of the content until the team was in agreement. This process resulted in a prototype intervention that addressed key barriers and facilitators identified in the primary qualitative work and metasynthesis and that fit with the guiding principles, behavioral analysis, and logic model.

The content was transferred into web-based pages in LifeGuide (a software program for developing web-based interventions [21]), and further discussions and iterations took place, modifying the presentation of information, before reaching an agreement on the prototype to be used in the intervention optimization phase.

Phase 2: Intervention Optimization

Design

Following the PBA, think-aloud qualitative studies were used to optimize a digital prototype of an intervention [15]. They are designed to explore users' perspectives on the intervention's content and its relevance, appropriateness, and understanding as well as users' views on functionality, usability, and design of an intervention. Ethical approval for the study was granted by the National Health Service South Central Oxford B Research Ethics Committee. Participants provided informed consent and were identified only by participant ID.

Participants

Participants were health professionals from primary care practices in the south of England who responded to a call from the local National Institute for Health Research (NIHR) clinical research network Wessex for health practitioners to participate in a semistructured face-to-face interview. The study team then contacted interested participants and arranged to meet them at their home or at their practice.

Procedure

Eligible participants met with researchers (HB, TK, RL, and/or SW) who invited them to engage with the digital prototype intervention and say what they were thinking aloud. When necessary, the interviewer prompted participants (eg, asking practitioners, "How do you feel about the information on this page?"). Interviews ranged from 32 to 82 minutes in length and were audio recorded. The interview schedule is presented in Multimedia Appendix 1.

Following each interview, comments were discussed among the study team. Comments were initially discussed based on notes from the interviews, and further discussions were conducted based on the analyzed transcriptions. This was done to modify the intervention in a timely manner. As a result of these discussions, amendments were made to the intervention in an iterative fashion; 4 subsequent intervention versions were shown to the participants following the first prototype. Participants participated in only 1 interview. Therefore, participants who were recruited later in the study saw an updated and amended version of the intervention, different from the versions seen by participants recruited earlier. This process allowed the changes that were made as a result of practitioner feedback to be shown and discussed with practitioners recruited later in the study. Interviews with practitioners continued until data saturation was reached, that is, when comments about the intervention reflected that no further changes were necessary and there were, therefore, no new codes identified.

Analysis

Interviews were transcribed verbatim and analyzed using 2 analytical methods. The first method comes from the PBA and is more rapid than thematic analysis, involving the use of coding tables specifically designed to highlight positive and negative comments about the intervention. All minor amendments to the intervention were agreed upon following discussion with the intervention development team. Issues that were likely to affect participant engagement or intervention effectiveness were

XSL•FO
RenderX

discussed with the broader group before making changes. Along with positive and negative comment tabulation, we also conducted a thematic analysis of general or overall views about the intervention using NVivo software (QSR International) [22,23]. This more in-depth analysis aimed to capture participants' views of the intervention and ideas about how they may engage with it. The thematic analysis was central in making amendments to the intervention. HB independently coded the transcripts using thematic analysis and discussed an initial coding frame with a second researcher (AG). The broader team was included in discussions about labeling and interpretation. The findings of the in-depth thematic analysis are presented here.

Results

Phase 1: Intervention Planning and Development

Step 1: Systematic Reviewing

Our qualitative thematic synthesis identified key barriers and facilitators to discontinuation for both patients and health professionals (refer to the study by Maund et al [13] for full details). The following issues were regarded as key to inform the development of the health professional intervention:

- GPs reported patient dependence on antidepressants and patients' difficult life circumstances as barriers to discontinuing antidepressants.
- Patients perceived their doctor as the navigator of discontinuation, responsible for recommending or approving the decision to discontinue, and responsible for initiating the discussion.
- Patients expected their doctor to support and guide them through the discontinuation process.
- Both patients and GPs acknowledged that the GP often does not have the time necessary to best support discontinuation.

The findings from this thematic synthesis informed the guiding principles, behavioral analysis, and logic model, which formed the basis for intervention content selection and development.

Step 2: Primary Qualitative Research

Thematic analysis [22] of the focus groups and interviews resulted in 5 themes exploring barriers and facilitators for health professionals initiating and discussing discontinuation with patients (refer to the study by Bowers et al [14] for further details of the findings). After discussion, the following issues were determined to be the most relevant and directly informed intervention development:

- When neither the health professional nor the patient raises the topic of discontinuation, this may result in a form of collusion where both parties are assuming that the other party wants antidepressant treatment to continue.
- Health professionals were concerned about destabilizing well patients—some felt that continuing antidepressants was easier and less risky.
- Knowing the patient's history and past experiences with depression and antidepressants facilitated discontinuation discussions, although poor continuity made such conversations less likely.
- Insufficient information and understanding of antidepressant discontinuation made them less confident in their ability to successfully taper patients off antidepressants.
- Conversations about discontinuation are time consuming and require follow-up appointments, which are often limited.

These literature syntheses and qualitative interview findings informed the development of guiding principles, a behavioral analysis, the logic model, and the intervention content. For example, understanding that the GP's time constraints can be a barrier resulted in a key design objective being *to provide rapidly accessible information designed to increase self-efficacy for effectively managing discontinuation in patients* (Textbox 1). The qualitative findings also directly informed the content of the intervention. For example, recognizing the collusion that results in both parties assuming that the other wants treatment to continue led to an intervention module dedicated to *broaching the subject* (Textbox 2).

XSL•FO
RenderX

Textbox 2. Outline of the intervention content.

Content and Description

- Why reduce
  - A rationale for discontinuation highlighting patients' views on antidepressants and why they would prefer not to take them and covering evidence on relapse rates for patients discontinuing in primary care
- Broaching the subject
  - Guidance on addressing the topic of discontinuation within a consultation and acknowledging time constraints
- When to start tapering
  - Information about which patients might be appropriately considered for discontinuation and what time might be most appropriate to initiate the withdrawal
- Reduction schedules
  - Suggested reduction schedules for different medications under different circumstances (eg, depending on drug half-life or patients' previous experiences with withdrawal)
- Dealing with withdrawal symptoms
  - Information about distinguishing relapse from withdrawal symptoms and how to respond to mild, moderate, and severe withdrawal symptoms
- Dealing with relapse
  - Information about how patients are advised to manage their relapse prevention
  - Guidance on when it might be advisable to reinstate antidepressants in the face of relapse
- ADvisor for Health Professionals for patients
  - A brief overview of the content covered in the patient digital intervention to enable practitioners to recommend information in the web-based resource for patients to look at home
- Printable pages
  - Printable handouts to give to patients (eg, a tapering regime template)
- Resources
  - Links to relevant papers and guidelines related to antidepressant discontinuation

Step 3: Guiding Principles for the REDUCE Intervention Package

These principles included design objectives and design features (Textbox 1). Our 2 broad design objectives were (1) to build confidence that discontinuing antidepressant medication is safe and achievable in the long term and (2) to be an accessible, motivational resource that supports patients in managing their withdrawal, which aligns with their preferences. These objectives were achieved through the development of the patient digital intervention (which is described elsewhere [24]) and through the health professional intervention (eg, patients' views on enabling empathy and GP support informed the section for GPs to build their confidence that discontinuation is safe). The design features that support these objectives are presented in Textbox 1.

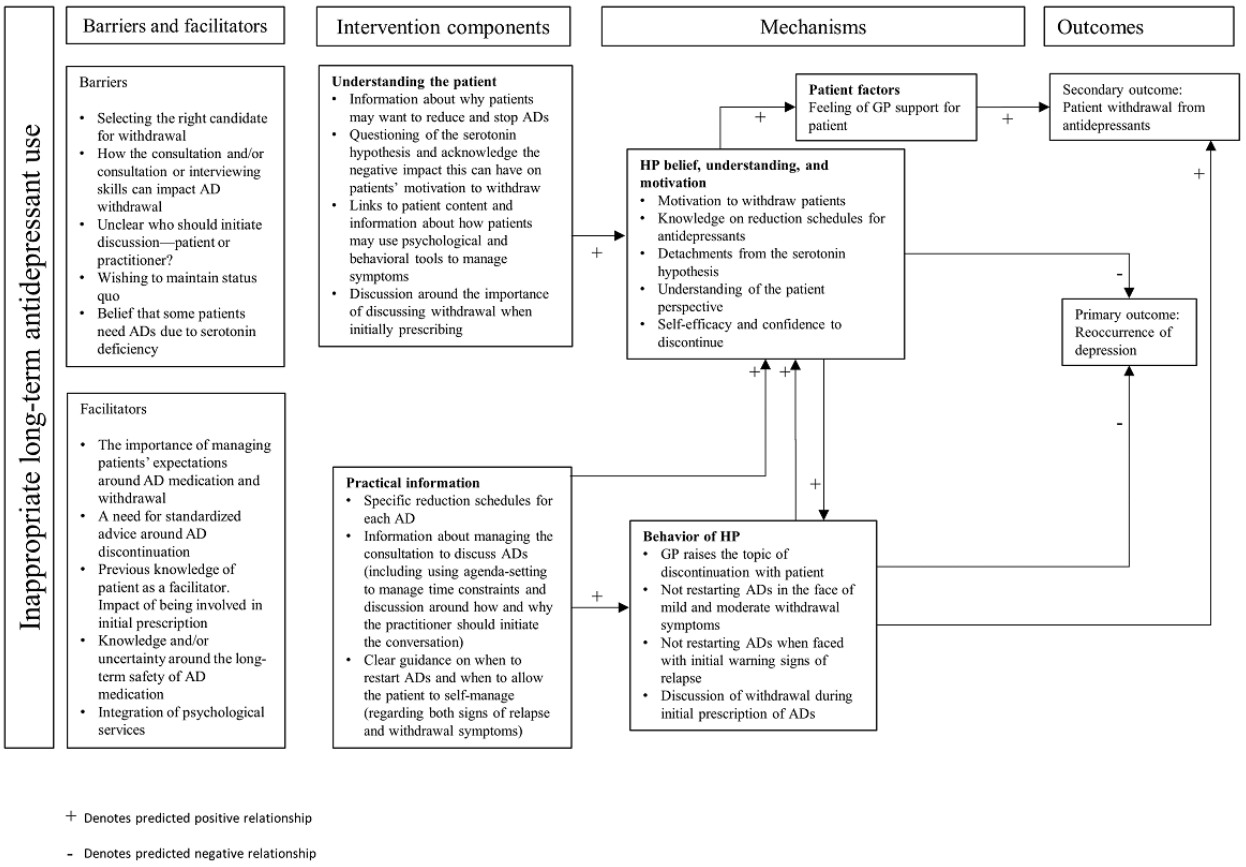
Step 4: Behavioral Analysis and Logic Modeling

The behavioral diagnosis for the health professional intervention can be found in Multimedia Appendix 2 [17,18]. Reducing and stopping antidepressant medication was our target behavior, and reflective motivation and psychological capability were considered the key constructs for changing this behavior. In terms of increasing the psychological capability of health professionals, this would involve improving knowledge about discontinuation and setting expectations around the withdrawal process to build self-efficacy in discontinuing antidepressants. To improve reflective motivation for health professionals, the intervention targeted beliefs about treatment that may act as barriers to discontinuation (eg, that withdrawal is always challenging and unachievable).

Content for particular intervention modules was drafted and mapped against constructs from the BCW and NPT (Multimedia Appendix 3). For example, in the module dedicated to providing a rationale for discontinuing antidepressants, the key BCW construct addressed is reflective motivation, and for NPT, the key constructs are enrollment and systematization. The logic model for the intervention is shown in Figure 2. Depicted in the logic model are the hypothesized relationships between barriers and facilitators to discontinuation intervention components, mechanisms, and outcomes.

XSL•FO
RenderX

Figure 2. Logic model for Advisor HP. AD: antidepressant; GP: general practitioner; HP: health practitioner.



Step 5: Drafting Intervention Content

A prototype digital web-based intervention was developed for health professionals involved in supporting patients to discontinue long-term antidepressant use. The contents of the web-based intervention are described in Textbox 2. A digital intervention for patients was also developed, as reported separately [24].

The intervention was designed to be flexible to cater to the time constraints under which GPs work. As such, the topics covered were optional and accessible from the main menu page; users could skip content that they felt would not be beneficial to them. This also meant that necessary information could be accessed quickly during a consultation.

Phase 2: Intervention Optimization

Participants

Of the 32 practitioners who expressed interest, 13 could not be reached by the study team or could not take part (eg, because of time constraints or sickness), resulting in a final sample of 19 practitioners (2 NPs and 17 GPs); there were 8 men and 11 women. The mean age was 44.29 (SD 7.13) years, and the mean number of years since qualification was 14.39 (SD 7.77) years (Table 1).

XSL•FO
RenderX

Table 1. Demographic information about think-aloud interview participants.

ID	Sex	Role	Years qualified	Iteration seen
GP[a]/15/01	m[b]	GP	4	1
GP/11/01	f[c]	GP	22	1
GP/11/02	f	GP	32	1
GP/13/01	m	GP	14	3
GP/14/01	m	GP	34	1
GP/16/01	f	GP	14	2
GP/17/01	f	GP	27	2
GP/18/01	f	GP	26	2
GP/08/01	m	GP	20	1
GP/19/01	m	GP	21	3
NP[d]/08/01	f	NP	17	1
GP/21/01	f	GP	5	4
GP/22/01	f	GP	16	4
GP/21/02	f	GP	8	5
GP/22/02	m	GP	Not recorded	4
NP/23/01	f	NP	Not recorded	5
GP/25/01	m	GP	12	4
GP/27/01	m	GP	23	5
GP/27/03	f	GP	13	5

[a]GP: general practitioner.

[b]m: male

[c]f: female.

[d]NP: nurse practitioner.

Findings

Summary of Thematic Analysis

The thematic analysis of the 19 transcripts resulted in the following five themes: how Advisor for Health Professionals (ADvisor HP) would be used in practice, pitching at the right level for health professionals, *ADvisor HP is evidence based*, the need for brevity and quick access, and *ADvisor HP is useful*. Participant identifiers are presented after each quote, and the iteration number refers to the iteration of the intervention that the practitioner saw (ie, iteration 1 means that the participant saw the first iteration of the intervention). The changes made at each iteration are listed in Textbox 3.

XSL•FO
RenderX

Textbox 3. Examples of changes made at each iteration based on the feedback from the think-aloud interview.

Changes made following iteration 1 feedback

- Clarified reporting of statistics on relapse rates after discontinuation
- Clarified that side effects improve once the antidepressant is stopped
- Provided further clarity around antidepressant use and discontinuation for treating anxiety
- Added qualifying statements such as "experienced GPs will know" to address comments around some of the information being well-known to more experienced general practitioners (GPs) but necessary for newly qualified GPs and nurse practitioners
- Added that there is evidence to support that asking about additional concerns in a consultation does not extend the length of a consultation
- Amended the definition of patients who are currently well to include those who may still have mild symptoms
- Emphasized that the included reduction schedules are suggestions and may be individualized
- Fixed typos
- Added clarity on the timeframe of withdrawal symptoms compared with symptoms of relapse
- Provided specific information about anxiety as a withdrawal symptom

Changes made following iteration 2 feedback

- Revised text to be more concise, using more bullet points
- Changed navigation, so that each module provides an overview with an option to read about it in more detail
- Revised tone to be less prescriptive

Changes made following iteration 3 feedback

- Formatting changes (eg, font choices, text box color, and image layout)
- Navigation changes (eg, making the end of a section and navigation back to the main menu clearer)
- Correction of typos

Changes made following iteration 4 feedback

- Navigation buttons amended (ie, removing "exit" buttons to help people navigate back to the main menu)
- Added clarity around what information patients get from their digital intervention
- Revised wording around the timing of tapering regimes

How ADvisor HP Would Be Used in Practice

Participants described using ADvisor HP as a resource of information which they can refer to. Nine participants felt that they would first read through ADvisor HP as a whole and that once all of the information was read, they would not need to refer back to the same information many times:

> *I really like it. Is it potentially going to be — I mean — not quite a one-off visit...but, you know, if you had worked all the way through it, yes, you don't have a photographic memory, but — because you've given us some nice, simplistic regimes, it isn't something necessarily that I'm going to refer to umpteen times a day [I: Right], because, of course, once you've done it three or four times, you're going to know what that regime is.* [GP/22/01, iteration 4]

Many health professionals liked that they could engage with the intervention in different ways, for example, *dipping in and out* (n=4) and using in a targeted fashion for specific knowledge, as needed (n=7). Two participants said they would save a shortcut on their computer for easy and quick access to information if and when they needed it:

> *I do like that you pick what you want to read and so you can dip in with what you think you want to know and dip out. It's your choice what you do; you don't have to read through the whole thing to get to what you're particularly interested in, because there will be times when you think — actually I just don't know what to do about — about that, so you can dip in and look at that. So, yes, I like that.* [GP/11/02, iteration 1]

Three health professionals (2 nurses and 1 GP) talked about using ADvisor HP more regularly and making it a part of everyday practice:

> *Yes, I would definitely use it in practice — on a day-to-day — like I said, just looking in between consultations or if I had a particular patient that I'd discussed it with and — I may even kind of ring them back if I'd looked at it after the consultation and make suggestions.* [NP/08/01, iteration 1]

Some health professionals talked about printing copies of the information both for easy access for themselves and to hand to patients, as they felt there was something beneficial in the act of physically giving something. Therefore, we added 2 printouts

XSL•FO
RenderX

to be given to patients—one to direct patients to specific content in the patient intervention, ADvisor, and another to be used to provide a personalized tapering regime:

> *Patients like to go out the room holding something, don't they, whether it's a prescription or something and to give them something to do for themselves, gives them a bit more ownership of it and it feels like they're doing something for themselves, rather than having to rely on us and the tablets all the time.* [GP/18/01, iteration 2]

Very few health professionals felt that they would use ADvisor HP within a consultation for their own information. However, some talked about using ADvisor HP collaboratively with a patient within a consultation as a way of showing the patient reliable information and providing evidence for their clinical decision making:

> *I think sometimes, although it's for the doctor, sometimes it's quite useful to show the patients these things as well. I don't tend to use the internet that much but sometimes when you get a leaflet or you find something that's kind of - and to us, if you show it to the patients as well, then it builds up that trust a bit more as well, that that's why you're doing it. You show evidence for things.* [GP/18/01, iteration 2]

When discussing how they envisage ADvisor HP being used in practice, 3 GPs reported that they did not feel they would use it. The reasons they provided were that it would be difficult to find the time to read it and that they already knew the information contained in ADvisor HP:

> *I think most GPs would love to get their patients off antidepressants; if we could give patients a link to a website supporting that, then that's great and we could come to a plan together, but I don't know that I necessarily need to read that, to make that plan.* [GP/16/01, iteration 2]

There was acknowledgment among some health professionals that work would need to be done to get other HPs to engage with and be aware of ADvisor HP:

> *I think it would be really helpful. I'm not quite sure how you're going to — market it as such, because obviously all of us get absolutely inundated with — there's a new tool for this and a new resource for that and new guidance on this and — we have to cherry pick what we actually even — bother looking at, let alone going through.* [GP/22/01, iteration 4]

Pitching It at the Right Level for GPs

Later iterations reduced the detail of information where health professionals highlighted that it was common knowledge or unnecessary and it was stated in ADvisor HP that information may not be new but may act as a reminder (Textbox 3). Two GPs also commented that the tone of the intervention could be less prescriptive:

> *So in, under this talking to the patient heading, sounds like a lot of it is so obvious that it's patronising and could be annoying to a GP.* [GP/15/01, iteration 1]

Some of these health professionals explained that it was still useful to include information they already knew, as it can act as a reminder and may be useful to less experienced practitioners and trainees. A GP explained that it is reassuring to read that the guidance corresponds to what they are already doing in practice:

> *I suppose a lot of it — you know — I feel I know already, in truth, but there's — it's a reminder. And actually, you know, it's a prompt that when — when I'm next speaking to somebody, I might just — and we're talking about coming off antidepressants, I might actually just specifically touch on those areas, whereas before I would have probably waited until they told me those things.* [GP/27/01, iteration 5]

A GP explained that information that may seem obvious to some GPs does not need to be excluded from ADvisor HP. By having a main menu and easy access to different sections, one could quickly move from one section to another if they felt that the information was not something they needed to read:

> *If you go on to a section and you think, I know that already, you could always come out of it.* [GP/11/02, iteration 1]

Despite these comments, many of the GPs reported learning something new in ADvisor HP, in particular, new information about antidepressants (eg, switching to another antidepressant to help avoid distressing withdrawal), patient perspectives (eg, patients expect their GP to initiate discontinuation conversations), and how psychological therapeutic techniques may help patients to discontinue antidepressants:

> *That's interesting. It says many patients may continue taking their antidepressant because they believe the GP wants them to, because I've always thought it was almost the opposite way around, that patients want to carry on their antidepressants and GPs are trying to take them off it. And I often find that people are saying — please don't make me stop them and wanting to carry on with them for longer, but probably I'm wrong; maybe that's just an anecdotal thing for me and it seems, from this, that they've shown that if a GP doesn't suggest discontinuing, that they'll just carry on and on and on.* [GP/27/03, iteration 5]

Two GPs explicitly stated that they would change their practice following new information in ADvisor HP. They would now be more inclined to broach the subject of stopping, more likely to ask patients about upcoming life events that may interfere and be less inclined to restart antidepressants in the face of withdrawal symptoms:

> *So if that [symptoms of depression or anxiety] happens within a couple of days of the dose change then it's probably not a relapse; it's probably more a withdrawal. But it's a good reminder for me, because I think if I had a patient come back to me, having cut their SSRI, then within a few days was saying they felt awful, I might just be tempted to put them back on it, really. [I: Right] Whereas actually*

XSL•FO
RenderX

> *now I might be tempted to — encourage them to stay with it for a bit longer.* [GP/27/01, iteration 5]

ADvisor HP is Evidence Based

Health professionals reported reassurance that the information was in line with the National Institute for Health and Care Excellence (NICE) guidance and liked the references to research to support the information in ADvisor HP:

> *So I guess it's reassuring to have some evidence-base and some NICE guidelines incorporated. As this is a practitioner website, I guess it — it's helpful to give peace of mind and reassure people that actually it's the right thing to do, to come off when — when certain criteria have been fulfilled, as it were* [GP/08/01, iteration 1]

In particular, evidence showing the relatively low relapse rates of patients was found to be reassuring by 5 of the GPs in the sample. They also felt that these figures would be useful in reassuring the patient, who may fear discontinuation:

> *I actually — I'm quite impressed with the number that came off; that's more than I thought. [I: Okay] So that's quite good and it probably would make me a bit more motivated to try and get people off.* [GP/16/02, iteration 2]

Also, 2 GPs acknowledged that the existing guidelines are not easy to navigate when looking for particular pieces of information or guidance. They felt that ADvisor HP made it easy to access the information needed:

> *So I think a lot of people would be interested in the research evidence for it. Yes, just looking who is eligible to discontinue, which I think is quite important, because NICE guidelines tend to be a little bit thick and long winded and it's hard to really condense them, I think, to get the information you need, sometimes, so that's quite helpful, having just those several bullets points there.* [GP/25/01, iteration 4]

The Need for Brevity

Many health professionals highlighted their strict time constraints and, thus, the need to access information quickly. Five GPs commented, particularly in the earlier iterations, that ADvisor HP was time consuming to go through and that this may prevent them from being able to use the intervention. A large number of health professionals reported that there was more information than necessary. Some suggested that although the content was useful, it could be presented more concisely:

> *I think the overwhelming thing is probably we have such limited time and we need to be able to look at everything so quickly and you need to be able to find something straightaway. So I think it is, because you very much go straight into it [the main menu] and then you go to Reduction Schedules, as long as you've looked at it before and know where they are.* [GP/21/01, iteration 4]

However, 4 GPs and 1 NP reported that the intervention was concise and not too big to go through. Nevertheless, in further iterations of ADvisor HP, the content was revised to be more concise. It is possible that these changes resulted in health professionals finding the content more manageable to go through:

> *I like the type font, I think it's quite clear, the photos are appropriate. It's very easy to read. The whole package isn't too big and you know what you're dealing with.* [GP/21/02, iteration 5]

ADvisor HP Is Useful

Throughout the interviews, GPs and nurses commented that the information in ADvisor HP was useful. There were mixed views, in that different pieces of information were reported as useful by different participants. This fits with the guiding principles of ADvisor HP, in that the information should be flexible so that participants can access what they find useful, without the need to view the information they may not wish to view at that time. Tapering regimes were highlighted as particularly useful by 13 of the 15 health professionals, which fits with our primary qualitative evidence that there is a lack of existing clear, accessible guidance:

> *I think the most helpful bits are probably actually looking at the advice on the medications and the exact dosage regimes to use and I think probably the information for the patients is all quite good.* [GP/21/01, iteration 4]

Two GPs did not find tapering guidance useful, reporting that they already knew it and that drug regimens need tailoring to each individual patient.

Many health professionals also reported that the information on how to deal with withdrawal symptoms and how to distinguish between them and signs of relapse was useful:

> *Yes, the contrast between withdrawal and relapse, especially the timeframe.... Yes, I think that's quite a good table; I like that* [GP/16/01, iteration 2]

By the end of the interviews, the main issues with ADvisor HP were resolved. The intervention was deemed ready for the forthcoming feasibility trial (which will be reported elsewhere).

Discussion

Principal Findings

Using the PBA, we developed a digital intervention designed to support health care professionals helping patients to withdraw from long-term antidepressant treatment. This intervention was based on a systematic literature review, in-depth qualitative research, and discussion with experts. The intervention aims to provide easily accessible, evidence-based guidance for health professionals and is grounded in practitioner perspectives and preferences. It was designed to be part of an intervention package along with a patient digital resource and patient telephone support with a psychological practitioner.

The think-aloud interview findings suggest that ADvisor HP could work in practice, as the information is quickly accessible and based on the latest evidence. The intervention will be examined in a feasibility randomized control trial in which a

XSL•FO
RenderX

qualitative and quantitative process evaluation will be carried out. This feasibility study will explore how practitioners engage with the intervention to support to patients discontinuing antidepressants. Further intervention modifications may occur in response to the findings of the feasibility trial, in line with the latter stages of the PBA [15].

A digital intervention may be particularly useful in imparting knowledge and possibly more limited in changing practitioners' attitudes and developing their skills. Educational research suggests that attitudes toward mental health problems may need to be challenged in peer group discussions and that a combination of modeling, role-play, and video feedback might be needed for the acquisition of new skills [25]. However, such educational initiatives are expensive and time consuming, and given the high prevalence of inappropriate long-term prescription of antidepressants and the time pressures under which primary care practitioners work, interventions that can quickly be applied at scale are needed.

Many previously developed tools to support practitioners in deprescribing have focused on clinical decision-making support systems or polypharmacy, particularly in older adult patients [26,27]. These tools often result in increased deprescribing; however, our qualitative work suggests that further education and belief or attitude change around antidepressant discontinuation would be needed to support deprescribing for long-term antidepressants. A study of antibiotic prescribing provided GPs with an internet education tool that was optimized using think-aloud methods [28]. This intervention resulted in significantly lower rates of antibiotic prescribing [29]. In using the PBA to develop ADvisor HP, health professionals' beliefs have informed the initial development and the optimization of the intervention, which may therefore support belief and behavior change.

Limitations

This study has some limitations. Many of the practitioners recruited for the focus groups and think-aloud interviews were recruited from less deprived regions of the south of England. Many patients who live in these regions are from more affluent backgrounds, with few from ethnic minorities. Therefore, the perspectives in this work may not represent broader cultures and experiences of antidepressant treatment nationally. In the feasibility work and the subsequent fully powered randomized control trial, further qualitative work will be carried out with a larger and more diverse population of practitioners in England and Wales.

It is possible that the practitioners who were interested in participating were already interested in mental health research and may, therefore, be more knowledgeable than other practitioners who may show a less keen interest in this research topic. The potential bias in our qualitative sample may explain why some GPs felt that the information was not new. Some GPs who participated in the think-aloud interviews felt that they did not need the information in the intervention, but nevertheless, they thought that it would be useful for GP trainees and doctors who are relatively new to UK family practice. Recruiting practices in different regions with a more diverse sample in the feasibility and main trials may encourage participating practitioners who are less informed about antidepressant discontinuation to take part. This development work included only 2 NPs, of whom both were prescribers responsible for managing long-term antidepressant use in primary care. Many practices do not involve NPs in the management and discontinuation of antidepressants. This limited representation of NPs made it difficult to identify differences between GPs' perspectives and nurses' perspectives. Conducting more research with NPs and other health care professionals who may be involved in antidepressant discontinuation would be useful to further understand how this intervention may be used. Exploring views on the intervention with a broader sample will help ensure that the intervention is pitched at the right level.

Conclusions

This development work has demonstrated the need for a brief and accessible intervention for health professionals. In addition, the think-aloud interviews suggested that the developed intervention may contribute to a change in attitudes and behaviors related to antidepressant discontinuation. The randomized trial will test whether the intervention works in terms of enhancing the discontinuation of antidepressants, while measuring the effects on depressive symptoms.

Acknowledgments

The ADvisor HP intervention was developed using LifeGuide software and/or methodologies, which was partly funded by the NIHR Southampton Biomedical Research Centre. The authors would like to acknowledge Emma Maund who conducted the 2 systematic reviews that informed the intervention development.

Conflicts of Interest

TK reports grants from the NIHR during the conduct of the study. JM reports grants from the NIHR during the conduct of the study and is a board member of the Council for Evidence-Based Psychiatry, an unfunded organization, whose mission is to "communicate evidence of the potentially harmful effects of psychiatric drugs to the people and institutions in the UK that can make a difference," and a cochairperson of the Critical Psychiatry Network. CD reports grants from the NIHR during the conduct of the study. MM reports grants from the NIHR during the conduct of the study.

Multimedia Appendix 1
Interview schedule.

XSL•FO
RenderX

[DOCX File , 61 KB - jmir_v23i7e25537_app1.docx]

Multimedia Appendix 2
Behavioral diagnosis.
[DOCX File , 18 KB - jmir_v23i7e25537_app2.docx]

Multimedia Appendix 3
Theoretical modeling for intervention content.
[DOCX File , 25 KB - jmir_v23i7e25537_app3.docx]

References

1. Prescription Cost Analysis - England, 2018 [PAS]. Prescription Cost Analysis, NHS Digital. 2019. URL: https://digital.nhs.uk/data-and-information/publications/statistical/prescription-cost-analysis/2018 [accessed 2021-06-07]
2. Taylor S, Annand F, Burkinshaw P, Greaves F, Kelleher M, Knight J, et al. Dependence and withdrawal associated with some prescribed medicines : an evidence review. Public Health England, London. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940255/PHE_PMR_report_Dec2020.pdf [accessed 2021-06-21]
3. Kendrick T, Stuart B, Newell C, Geraghty AW, Moore M. Did NICE guidelines and the Quality Outcomes Framework change GP antidepressant prescribing in England? Observational study with time trend analyses 2003-2013. J Affect Disord 2015 Nov 01;186:171-177. [doi: 10.1016/j.jad.2015.06.052] [Medline: 26241666]
4. Moore M, Yuen HM, Dunn N, Mullee MA, Maskell J, Kendrick T. Explaining the rise in antidepressant prescribing: a descriptive study using the general practice research database. Br Med J 2009 Oct 15;339:b3999 [FREE Full text] [doi: 10.1136/bmj.b3999] [Medline: 19833707]
5. Ambresin G, Palmer V, Densley K, Dowrick C, Gilchrist G, Gunn JM. What factors influence long-term antidepressant use in primary care? Findings from the Australian diamond cohort study. J Affect Disord 2015 May 01;176:125-132. [doi: 10.1016/j.jad.2015.01.055] [Medline: 25704565]
6. Cruickshank G, Macgillivray S, Bruce D, Mather A, Matthews K, Williams B. Cross-sectional survey of patients in receipt of long-term repeat prescriptions for antidepressant drugs in primary care. Ment Health Fam Med 2008 Jun;5(2):105-109 [FREE Full text] [Medline: 22477855]
7. Piek E, Kollen BJ, van der Meer K, Penninx BW, Nolen WA. Maintenance use of antidepressants in Dutch general practice: non-guideline concordant. PLoS One 2014 Feb;9(5):e97463-e97467 [FREE Full text] [doi: 10.1371/journal.pone.0097463] [Medline: 24858011]
8. Ferguson JM. SSRI antidepressant medications: adverse effects and tolerability. Prim Care Companion J Clin Psychiatry 2001 Feb;3(1):22-27 [FREE Full text] [doi: 10.4088/pcc.v03n0105] [Medline: 15014625]
9. McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. Paying the Price: The Cost of Mental Health Care in England to 2026. London: The King's Fund; Oct 2008:1-165.
10. Sinclair JE, Aucott LS, Lawton K, Reid IC, Cameron IM. The monitoring of longer term prescriptions of antidepressants: observational study in a primary care setting. Fam Pract 2014 Aug;31(4):419-426. [doi: 10.1093/fampra/cmu019] [Medline: 24850795]
11. Eveleigh R, Grutters J, Muskens E, Voshaar RO, van Weel C, Speckens A, et al. Cost-utility analysis of a treatment advice to discontinue inappropriate long-term antidepressant use in primary care. Fam Pract 2014 Oct;31(5):578-584. [doi: 10.1093/fampra/cmu043] [Medline: 25121977]
12. Johnson CF, Macdonald HJ, Atkinson P, Buchanan AI, Downes N, Dougall N. Reviewing long-term antidepressants can reduce drug burden: a prospective observational cohort study. Br J Gen Pract 2012 Nov;62(604):e773-e779 [FREE Full text] [doi: 10.3399/bjgp12X658304] [Medline: 23211181]
13. Maund E, Dewar-Haggart R, Williams S, Bowers H, Geraghty AW, Leydon G, et al. Barriers and facilitators to discontinuing antidepressant use: a systematic review and thematic synthesis. J Affect Disord 2019 Feb 15;245(1):38-62 [FREE Full text] [doi: 10.1016/j.jad.2018.10.107] [Medline: 30366236]
14. Bowers HM, Williams SJ, Geraghty AW, Maund E, O'brien W, Leydon G, et al. Helping people discontinue long-term antidepressants: views of health professionals in UK primary care. BMJ Open 2019 Jul 04;9(7):e027837 [FREE Full text] [doi: 10.1136/bmjopen-2018-027837] [Medline: 31278099]
15. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
16. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015 Apr 23;1(1):37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]

XSL•FO
RenderX

17. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
18. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. London: Silverback Publishing; 2014:1-329.
19. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554 [FREE Full text] [doi: 10.1177/0038038509103208]
20. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions. UK Medical Research Council (MRC) guidance. 2017. URL: https://mrc.ukri.org/documents/pdf/mrc-phsrn-process-evaluation-guidance-final/ [accessed 2021-06-21]
21. Williams S, Yardley L, Weal M, Wills G. Introduction to LifeGuide: Open-source Software for Creating Online Interventions for Health Care, Health Promotion and Training. In: Medetel: Global Telemedicine and eHealth Updates. Zurich, Switzerland: International Society for Telemedicine & eHealth; 2010:187-190.
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
23. Joffe H, Yardley L. Content and thematic analysis. In: Marks DF, Yardley L, editors. Research Methods for Clinical and Health Psychology. Thousand Oaks, California, United States: SAGE Publications Ltd; 2003:1-248.
24. Bowers HM, Kendrick T, Glowacka M, Williams S, Leydon G, May C, et al. Supporting antidepressant discontinuation: the development and optimisation of a digital intervention for patients in UK primary care using a theory, evidence and person-based approach. BMJ Open 2020 Mar 08;10(3):e032312 [FREE Full text] [doi: 10.1136/bmjopen-2019-032312] [Medline: 32152159]
25. Gask L, Rogers A, Oliver D, May C, Roland M. Qualitative study of patients' perceptions of the quality of care for depression in general practice. Br J Gen Pract 2003 Apr;53(489):278-283 [FREE Full text] [Medline: 12879827]
26. Reeve E. Deprescribing tools: a review of the types of tools available to aid deprescribing in clinical practice. J Pharm Pract Res 2020 Feb 16;50(1):98-107. [doi: 10.1002/jppr.1626]
27. Scott IA, Pillans PI, Barras M, Morris C. Using EMR-enabled computerized decision support systems to reduce prescribing of potentially inappropriate medications: a narrative review. Ther Adv Drug Saf 2018 Sep 16;9(9):559-573 [FREE Full text] [doi: 10.1177/2042098618784809] [Medline: 30181862]
28. Anthierens S, Tonkin-Crine S, Douglas E, Fernandez-Vandellos P, Krawczyk J, Llor C, GRACE INTRO study team. General practitioners' views on the acceptability and applicability of a web-based intervention to reduce antibiotic prescribing for acute cough in multiple European countries: a qualitative study prior to a randomised trial. BMC Fam Pract 2012 Oct 11;13(9):101-173 [FREE Full text] [doi: 10.1186/1471-2296-13-101] [Medline: 23110756]
29. Little P, Stuart B, Francis N, Douglas E, Tonkin-Crine S, Anthierens S, GRACE consortium. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. Lancet 2013 Oct 05;382(9899):1175-1182 [FREE Full text] [doi: 10.1016/S0140-6736(13)60994-0] [Medline: 23915885]

Abbreviations

ADvisor HP: Advisor for Health Professionals
BCW: Behavior Change Wheel
GP: general practitioner
NICE: National Institute for Health and Care Excellence
NIHR: National Institute of Health Research
NP: nurse practitioner
NPT: normalization process theory
PBA: person-based approach
REDUCE: Reviewing Long Term Antidepressant Use by Careful Monitoring in Everyday Practice

XSL•FO
RenderX

Edited by R Kukafka; submitted 04.01.21; peer-reviewed by A AL-Asadi, C Jacob, J Parsons; comments to author 26.02.21; revised version received 22.04.21; accepted 27.04.21; published 16.07.21.

Please cite as:
Bowers H, Kendrick T, van Ginneken N, Glowacka M, Williams S, Leydon GM, May C, Dowrick C, Moncrieff J, Johnson CF, Moore M, Laine R, Geraghty AWA
A Digital Intervention for Primary Care Practitioners to Support Antidepressant Discontinuation (Advisor for Health Professionals): Development Study
J Med Internet Res 2021;23(7):e25537
URL: https://www.jmir.org/2021/7/e25537
doi:10.2196/25537
PMID:34269688

©Hannah Bowers, Tony Kendrick, Nadja van Ginneken, Marta Glowacka, Samantha Williams, Geraldine M Leydon, Carl May, Christopher Dowrick, Joanna Moncrieff, Chris F Johnson, Michael Moore, Rebecca Laine, Adam W A Geraghty. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

The Efficacy of eHealth Interventions for the Treatment of Adults Diagnosed With Full or Subthreshold Binge Eating Disorder: Systematic Review and Meta-analysis

Elnaz Moghimi[1], PhD; Caroline Davis[1], PhD; Michael Rotondi[1], PhD

School of Kinesiology and Health Science, Faculty of Health, York University, Toronto, ON, Canada

Corresponding Author:
Elnaz Moghimi, PhD
School of Kinesiology and Health Science
Faculty of Health
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100 ext 77239
Email: elnazm@yorku.ca

Abstract

Background: There has been a recent rise in the use of eHealth treatments for a variety of psychological disorders, including eating disorders.

Objective: This meta-analysis of randomized controlled trials is the first to evaluate the efficacy of eHealth interventions specifically for the treatment of binge eating disorder (characterized by compulsive overconsumption of food, in a relatively short period, and without compensatory behaviors such as purging or fasting).

Methods: A search on the electronic databases PubMed, Web of Science, Embase, MEDLINE, and CINAHL was conducted for randomized controlled trials that compared the efficacy of eHealth treatment interventions with waitlist controls.

Results: From the databases searched, 3 studies (298 participants in total) met the inclusion criteria. All interventions were forms of internet-based guided cognitive behavioral therapy. The results of the analysis demonstrated that when compared with waitlist controls, individuals enrolled in eHealth interventions experienced a reduction in objective binge episodes (standardized mean difference [SMD] −0.77, 95% CI −1.38 to −0.16) and eating disorder psychopathology (SMD −0.71, 95% CI −1.20 to −0.22), which included shape (SMD −0.61, 95% CI −1.01 to −0.22) and weight concerns (SMD −0.91, 95% CI −1.33 to −0.48). There was no significant difference in BMI between the eHealth interventions and controls (SMD −0.01, 95% CI −0.40 to 0.39).

Conclusions: These findings provide promising results for the use of internet-based cognitive behavioral therapy for binge eating disorder treatment and support the need for future research to explore the efficacy of these eHealth interventions.

(J Med Internet Res 2021;23(7):e17874) doi:10.2196/17874

KEYWORDS

internet; cognitive behavioral therapy; guided self-help; obesity; weight loss; eating disorder; binge eating; mobile phone

Introduction

Background

Binge eating disorder (BED) is recognized in the Diagnostic and Statistical Manual of Mental Disorders as abnormal and excessive eating patterns marked by uncontrolled, recurrent, and persistent binge eating [1]. In the individual, binge eating evokes guilt and distress but without compensatory weight loss behaviors (eg, purging), such as those found in individuals with bulimia nervosa. It is the most commonly diagnosed eating disorder (ED) compared with anorexia and bulimia nervosa [2] and is estimated to have a global pooled prevalence of 0.9%, with prevalence rates higher in women (1.4%) than in men (0.4%) [3].

BED is of particular research interest because of its frequency in primary care, its comorbidity with obesity and other medical and psychiatric disorders, and its high socioeconomic impact as a result of reduced quality of life and an increased need for

XSL•FO
RenderX

patients to use health and medical services [4-6]. Individuals diagnosed with BED have higher inpatient, hospital-based outpatient, and prescription-medication utilization and expenditure compared with age- and sex-matched controls, both before and after their BED diagnosis [7]. The health care utilization found in patients with BED is comparable with that found in people with other EDs and major psychiatric disorders [8]. Therefore, it is imperative that effective treatments are available for these patients to help reduce health care costs and provide long-term benefits.

The current *gold standard*, evidence-based treatment for BED is cognitive behavioral therapy (CBT). CBT has been shown to reduce binge eating frequency, lead to mild weight reduction [9,10], and cause complete abstinence in 50%-60% of patients post treatment [11,12]. Despite efficacious therapies available for BED, there continues to be low rates of help seeking for this debilitating mental health condition [13]. The important reasons for these trends may be the personal feelings of shame and fear, ED-related beliefs and perceptions, and a lack of access or availability of the treatment [14,15]. Thus, it is important that novel therapies address these barriers to treatment seeking.

Recently, the use of eHealth technology has been proposed as a potentially effective alternative to traditional, in-person treatment delivery for those with BED [16-19]. Although the term *eHealth* has many facets, it is generally defined as the use of emerging information and communication technology, particularly via the internet, to improve or enable health and health care [20]. Examples include, but are not limited to, email; software programs; teleconferencing or web conferencing; digital and mobile communication; and computer, mobile, or internet apps. The benefit of eHealth treatments is that they can be administered easily, are more accessible (particularly for patient groups that do not live near urban centers), can be used anonymously, and may reduce feelings of shame and fear [15]. Furthermore, eHealth treatments may be more cost-effective because in-person treatments often require expensive resources and infrastructure [15,21,22]. Although the first internet-delivered psychological treatments only emerged in the late 1990s [23], reviews exploring the effectiveness of internet-delivered treatments for EDs have yielded positive results in terms of their impact on quality of life, binge eating, compliance, dropout, and related psychopathology [24-28].

Many eHealth treatments for BED are in line with the CBT principles described in the self-help book *Overcoming binge eating* [29]. Some studies have described this approach as internet-based cognitive behavioral therapy (I-CBT), whereas other studies have described this approach as internet-based guided self-help (I-GSH). Both therapies consist of a combination of web-based psychoeducation, writing assignments or modules, and self-monitoring such as daily eating and activity diaries [10,30-34]. In most studied cases, I-CBT and I-GSH are typically guided to varying degrees by therapists, who provide support and answer questions and concerns that the patient may have.

Objective

Although several studies have examined the effects of eHealth treatments on different elements of BED, including bulimia [10,24,30-35], the overall effectiveness of these treatments, for this unique clinical population, has yet to be explored. Therefore, the objective of this systematic review and meta-analysis is to determine the effectiveness of eHealth treatments in adults diagnosed with full or subthreshold BED. Specifically, important hallmarks of the disorder, including binge episodes, BMI, and ED psychopathology, will be explored.

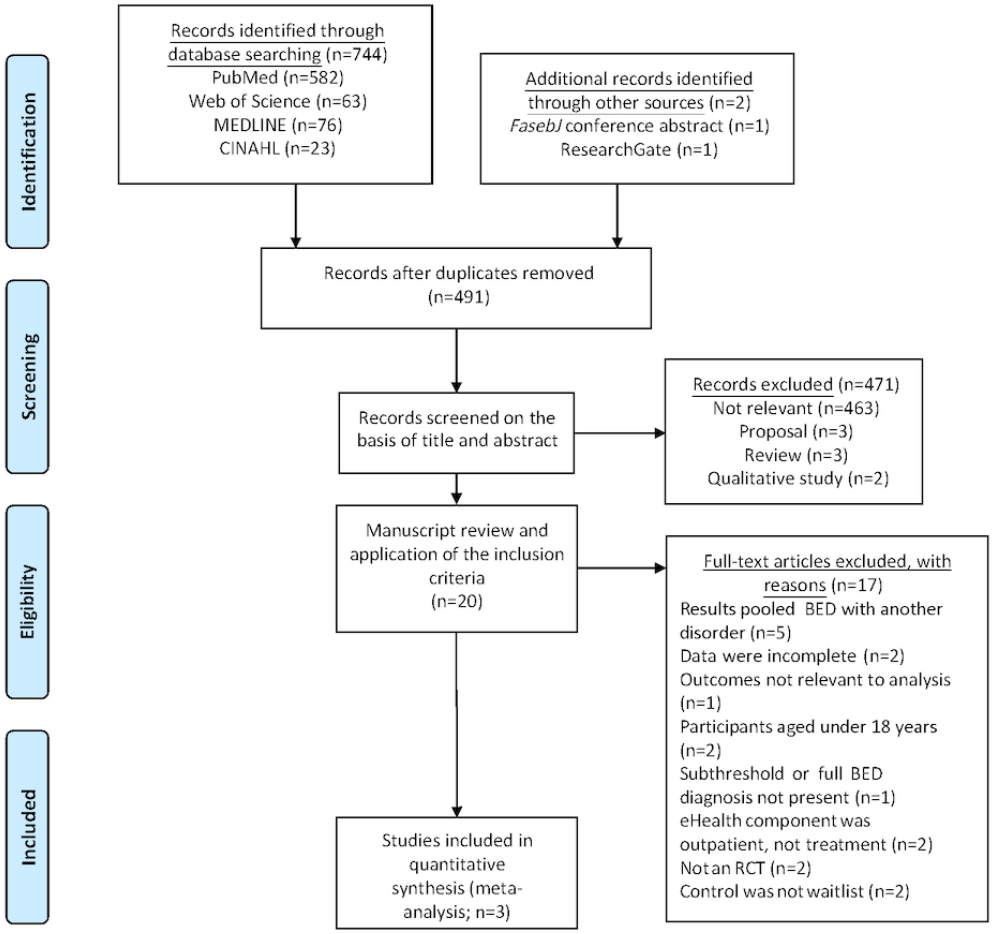
Methods

Search Strategy

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [36] (Multimedia Appendix 1) were followed when searching the electronic databases PubMed, Web of Science, Embase, MEDLINE, and CINAHL. A combination of MeSH (Medical Subject Headings) terms and keywords that represented the terms BED and eHealth were used to develop the algorithms. Examples of terms include *bing**, *binge eating*, *web-based treatment*, *internet*, *mobile phone*, *smartphone*, *telemedicine*, *telehealth*, and *remote*. Searches were performed until February 2019 and returned a total of 744 results. Articles were restricted to English and randomized controlled trials (RCTs), where possible. In addition, 2 articles were found via *FasebJ* and ResearchGate. A detailed outline of the literature search is presented in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart outlining the selection of studies included in the meta-analysis. BED: binge eating disorder; RCT: randomized controlled trial.



Inclusion Criteria

Studies used in the analysis had to meet the following prespecified inclusion criteria: (1) adult male or female participants aged 18 years and older, diagnosed with full or subthreshold BED. In the latter case, patients had to meet the criterion for objective binge episodes (OBEs) but could lack one of the other Diagnostic and Statistical Manual of Mental Disorders criteria (ie, frequency of less than 2 days with OBEs in 6 months, no marked distress, or presence of only 2 of the 5 associated criteria) [1,33,37] (both full and subthreshold BED participants were included, as a recent study [38] showed that individuals with these conditions do not differ significantly on measures of weight and shape concerns, restraint, psychiatric distress, and history of seeking treatment for an eating or weight problem [38]); (2) the intervention was a form of eHealth treatment and was the main form of treatment and was not administered posttreatment as a form of outpatient care; (3) outcomes observed included at least one of the following: OBE, BMI, Eating Disorder Examination Questionnaire (EDE-Q) total score, EDE-Q weight concern, and EDE-Q shape concern. The EDE-Q is a self-report questionnaire that measures ED severity and consists of 4 subscales: restraint, eating concern, weight concern, and shape concern [39]. This study explores total scores and relevant subscales that are related to the cognitive-affective aspect of body image (ie, shape and weight concern), as body image disturbances are common in this population [40]; (4) included studies must be RCTs, in which participants were randomly allocated to either a treatment or control group. This was to ensure that a robust methodology was used in each study and ensure that all estimates of effectiveness were not confounded by other factors and to enable the valid pooling of the data; (5) the comparison or control group consisted of a waitlist (WL) or no treatment; and (6) mean and SD pre- and posttreatment (or data available to calculate them) were reported. Studies that analyzed BED with other disorders were excluded if they did not stratify the ED of interest. If the eHealth component was outpatient, used subsequent to another main treatment, and not the main form of treatment, the study was excluded.

Data Extraction and Risk of Bias Assessment

The following data were extracted from each study: (1) first author's last name, (2) year of publication, (3) total sample size and group size posttreatment, (4) type of treatment and control, (5) BED diagnosis status and criteria used for diagnosis, (6) mean age of participants, (7) treatment length, (8) therapist contact, (9) percentage of females in the study, and (10) mean and SD posttreatment for all outcomes in the treatment and control groups.

The overall risk of bias was assessed by 2 independent reviewers using the revised Cochrane risk of bias tool for randomized

XSL•FO
RenderX

trials [41]. Any disagreements in the assessments were resolved via discussion. The following 5 domains were assessed: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in outcome measurement, and (5) bias in selection of the reported result. Each study was summarized as having a low, medium, or high risk of bias (Tables 1 and 2).

Table 1. Characteristics of the studies included in the meta-analysis. Waitlist controls were compared with an internet version of cognitive behavioral therapy or guided self-help therapy. All 3 therapies use the principles of cognitive behavioral therapy and have varying degrees of therapist guidance and interaction with the patients.

Authors	Participant			BED[a] diagnosis (diagnostic criteria)	Therapist contact	Females, n (%)	Age, mean (SD)	Treatment length	Risk of bias
	Total, N	Treatment group, n (%)	Control group, n (%)						
Carrard et al, 2011A [31]	74	37 (50) I-GSH[b]	37 (50) WL[c]	Full (n=43) or subthreshold (n=31) BED (eating disorder questionnaire based on DSM-IV[d])	Weekly contact via email	74 (100)	36 (11.4)	6 months	Low
ter Huurne et al, 2015 [30]	85	43 (51) I-CBT[e]	42 (49) WL	Diagnosed with BED (participant self-report based on DSM-IV)	Internet-based contact with the therapist twice a week	85 (100)	40.2 (11.4)	15 weeks	Low
Wagner et al, 2016 [10]	139	69 (50) I-CBT	70 (50) WL	Diagnosed with BED (telephone interview using DSM-5[f] criteria)	Internet-based contact with therapist when submitting assignments	134 (96.4)	35.1 (9.9)	16 weeks	Low

[a]BED: binge eating disorder.

[b]I-GSH: internet-guided self-help.

[c]WL: waitlist.

[d]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edition.

[e]I-CBT: internet-based cognitive behavioral therapy.

[f]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition.

Table 2. Postreatment results for each study.

Authors	OBE[a]		BMI		EDE-Q[b]		Shape concern		Weight concern	
	Treatment posttreatment, mean (SD)	Control posttreatment, mean (SD)	Treatment posttreatment, mean (SD)	Control posttreatment, mean (SD)	Treatment posttreatment, mean (SD)	Control posttreatment, mean (SD)	Treatment posttreatment, mean (SD)	Control posttreatment, mean (SD)	Treatment posttreatment, mean (SD)	Control posttreatment, mean (SD)
Carrard et al, 2011A [31]	5.5 (7.4)	9.1 (8.8)	29.2 (6.0)	27.9 (5.4)	2.5 (1.1)	2.9 (1.0)	3.7 (1.3)	4.1 (1.3)	__[c]	—
ter Huurne et al, 2015 [30]	—	—	—	—	2.6 (1.3)	3.2 (0.9)	3.5 (1.6)	4.2 (1.1)	3.1 (1.4)	3.9 (0.9)
Wagner et al, 2016 [10]	6.8 (7.5)	14.9 (7.7)	31.4 (6.9)	32.8 (8.3)	2.5 (1.2)	3.7 (0.8)	3.4 (1.4)	4.5 (0.8)	3.0 (1.3)	4.2 (0.8)

[a]OBE: objective binge episode.

[b]EDE-Q: Eating Disorder Examination Questionnaire.

[c]Not available (missing data).

Statistical Analysis

Given the anticipated heterogeneity, the studies were pooled using a meta-analytic random effects model. The reported *P* values were two-sided, with $P<.05$ considered statistically significant. As all outcomes measured were continuous, means and SD were used to calculate effect size, which was expressed as a standardized mean difference (SMD), corrected using Hedges *g* for a small sample size to reduce positive bias. The SMD was used to ensure that the scales used in the different studies were standardized. SMDs were classified based on the level of effect: less than 0.20 signified a very small effect, 0.20 signified a small effect, 0.50 signified a medium effect, and 0.80 signified a large effect. All studies used in the analysis reported mean and SD values. The I^2 statistic [42] was used to

XSL•FO
RenderX

quantify between-study heterogeneity. Here, the percentage of total variation in the estimates of the effect due to between-study heterogeneity is reported [43]. I^2 values above 25% indicated low heterogeneity, 50% indicated moderate heterogeneity, and above 75% indicated substantial heterogeneity [43]. The restricted maximum likelihood model estimator was used to measure the between-study variance, τ^2 [44]. Data were analyzed using the statistical software package RStudio Desktop version 1.1.463, metafor package [45].

Results

Study Characteristics

In total, 3 RCT studies met all the inclusion criteria and were included in the meta-analysis, with a total of 298 participants. All studies were consistent in the intervention and study participants and were appropriately combined in a meta-analysis. Assessment of the risk of bias scores indicated a low risk of bias in all 3 studies. Furthermore, all the studies recruited female participants, except 1 that sampled 96.4% (134/139) females [10]. The mean age of participants ranged from 35.1 (SD 9.9) to 40.2 (SD 11.4) years. In 2 studies, all participants were diagnosed with full BED. In one study [31], participants consisted of a combination of full (n=43) or subthreshold (n=31) BED. Furthermore, the method of BED diagnosis varied among the studies; in 2 studies, diagnosis was made using the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria, and in 1 study, the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) criteria were used. Among the studies, 2 administered I-CBT as the treatment, and the other study administered an I-GSH, which also used CBT principles. Although all 3 studies were based on CBT principles, they had varying degrees of therapist contact, which ranged from weekly to any time the participant submitted an assignment. None of the studies were unguided. All the therapists in the study were clinical psychologists, except one [30] where therapists either had a Bachelor of Science in nursing or social work or a Master of Science in psychology. The studies also varied in treatment length, ranging from 15 weeks to approximately 26 weeks (6 months). One of the studies did not measure OBE or BMI [30], and another study did not report EDE-Q weight concern [31]. The characteristics of the included studies are summarized in Tables 1 and 2.

Quantitative Analysis

Overview

A summary of the meta-analysis results for each included outcome is described below and summarized in Table 3. Forest plots for each outcome are presented in Multimedia Appendix 2 [10,30,31].

Table 3. Summary of findings for the randomized controlled trial studies.

Outcome	Studies, n (%)	Participants, N	Effect size; SMD[a] (95% CI)	Heterogeneity, I^{2b} (%)	*P* value
Objective binge episodes[c]	2 (67)	213	−0.77 (−1.38 to −0.16)	77	.01
BMI	2 (67)	213	−0.01 (−0.40 to 0.39)	50	.96
EDE-Q[d] total[c]	3 (100)	298	−0.71 (−1.20 to −0.22)	77	.005
Shape concern[c]	2 (67)	298	−0.61 (−1.01 to −0.22)	64	.002
Weight concern[c]	2 (67)	224	−0.91 (−1.33 to −0.48)	56	<.001

[a]SMD: standardized mean difference.

[b]I^2 values above 25% indicated low heterogeneity, 50% indicated moderate heterogeneity, and above 75% indicated substantial heterogeneity [43].

[c]$P<.05$.

[d]EDE-Q: Eating Disorder Examination Questionnaire.

OBE Result

In total, 2 of the studies involving 213 participants evaluated OBE. Among the studies, one demonstrated a significant reduction in OBE in the treatment group compared with the WL control (SMD −1.06, 95% CI −1.42 to −0.70). The pooled SMD was −0.77 (95% CI −1.38 to −0.16; Figure S1 of Multimedia Appendix 2 and Table 3), which showed a statistically significant effect.

BMI Result

In total, 2 of the studies involving 213 participants evaluated BMI. None of the studies demonstrated a significant change in BMI in the treatment group compared with the WL group. The pooled SMD was −0.01 (95% CI −0.40 to −0.39; Figure S2 of Multimedia Appendix 2 and Table 3).

EDE-Q Total Score

In total, 3 of the studies involving 298 participants evaluated the EDE-Q total score. Two of the studies demonstrated a significant reduction in EDE-Q scores in the treatment group compared with WL groups (SMD −1.17, 95% CI −1.53 to −0.81 and SMD −0.53, 95% CI −0.96 to −0.10). The pooled SMD was statistically significant with an estimate of −0.71 (95% CI −1.20 to −0.22; Figure S3 of Multimedia Appendix 2 and Table 3).

EDE-Q Shape Concern

In total, 3 of the studies involving 298 participants evaluated the EDE-Q subscale of shape concern. Two of the studies demonstrated a significant reduction in shape concern scores in the treatment group compared with WL groups (SMD −0.96, 95% CI −1.31 to −0.61 and SMD −0.50, 95% CI −0.94 to

XSL•FO
RenderX

−0.07). The pooled SMD was −0.61 (95% CI −1.01 to −0.22; Figure S4 of Multimedia Appendix 2 and Table 3).

EDE-Q Weight Concern

In total, 2 of the studies involving 224 participants evaluated the EDE-Q subscale of weight concern. Both studies demonstrated a significant reduction in shape concern scores in the treatment group compared with WL groups (SMD −1.11, 95% CI −1.47 to −0.75 and SMD −0.67, 95% CI −1.11 to −0.23). The pooled SMD is −0.91 (95% CI −1.33 to −0.48; Figure S5 of Multimedia Appendix 2 and Table 3).

Discussion

Principal Findings

This study reports the first meta-analysis of RCTs designed to assess the efficacy of eHealth treatments for individuals diagnosed with BED. Due to its specificity, 3 studies met the inclusion criteria and were included in the analysis. All of these used an internet-based form of guided CBT therapy, wherein the degree of therapist interaction varied depending on the nature of the intervention that was administered. Due to the novelty of eHealth innovations and our study objectives, it was important to evaluate efficacy by restricting to the RCT design. Despite the limited number of studies, statistically significant results demonstrated the effectiveness of internet-based CBT, in combination with GSH treatment, in reducing binge episodes, ED psychopathology, and shape and weight concerns. Although the efficacy of conventional CBT therapy has been well demonstrated [46-48], currently there is insufficient data to claim that internet-based CBT has an evidence base or effect size that is comparable with in-person CBT.

Despite the moderate effect of the treatment in reducing the number of OBEs, internet-based therapies did not appear to produce a significant change in BMI. Notably, the lack of substantial weight loss has long been considered one of the principal drawbacks of the current CBT therapy for BED. For instance, Peat et al [49] found no significant difference in BMI between therapist-led, partially therapist-led, and structured self-help CBT. Similarly, in a recent meta-analysis comparing pharmacological, psychological, and combined treatments, CBT was effective in reducing binge episodes, but its effect on weight loss was minimal [50]. Furthermore, the impact on BMI may be indirect, via reduced binge frequency, and consists of an extended maintenance phase [51]. In a more recent study involving in-person CBT treatment, no significant change in BMI was observed between pre- and posttreatment assessments [52]. These results are also in line with a study comparing in-person and eHealth CBT treatments, where BMI did not decrease in either treatment group, nor were they significantly different between the two groups [33]. Taken together, these results indicate that regardless of how CBT is administered, its effects on weight loss are minimal to nonexistent.

In patients with BED, the purpose of CBT is to reduce binge eating frequency and body image dissatisfaction by altering destructive behaviors and thinking patterns, particularly those that involve eating, weight and shape, and psychosocial functioning [53]. This is in line with the findings of the current meta-analysis, where a moderate effect of the treatment in reducing shape and weight concerns was observed. There is no direct treatment emphasis on diet or weight loss. Indeed, it has been observed that eating patterns may shift from binging to less compulsive overeating, which does not have the same elements of guilt and compulsion associated with it. In this way, the caloric deficits may not be met, and consequently, BMI does not change. However, research has also found that a complete abstinence from binges is associated with significant improvements in dietary and psychological outcomes, which may improve weight status in the long term [54]. Although there was a significant reduction in binge episodes in this meta-analysis, none of the treatment studies included in this analysis resulted in complete abstinence following treatment. Interestingly, a recent study has shown that when compared with obese, non-BED participants, those diagnosed with BED have a significantly higher threshold for what comprises a *large amount of food* [55]. Furthermore, laboratory-based studies have also demonstrated that when compared with their non-BED counterparts, participants diagnosed with BED tend to have a significantly higher caloric intake and consume large amounts of food even during nonbinge-eating episodes [56,57].

In addition to exploring the efficacy of eHealth interventions, this meta-analysis highlights an important point regarding the status of eHealth and BED research. That is, despite its promising impact on improving BED symptomatology and its moderate effect in reducing ED psychopathology, there are only a small number of RCT studies that have evaluated the efficacy of eHealth treatments, and even fewer studies have compared them with the traditional, in-person method of delivery [33,34]. Specifically, only 2 other studies compared eHealth treatments with in-person therapy [33,34]. Although the efficacy of internet treatments was demonstrated in both studies, when compared with in-person treatment, one study reported inferiority [33] and the other found no significant differences between the 2 treatment types [34]. Although the results of these studies are promising, further research will provide a better understanding of the merits of internet versus in-person treatment. However, our work shows that internet treatment is efficacious and may represent a more accessible and cost-effective treatment option for those who have difficulty obtaining in-person treatment for their ED. Given that BED is the most common ED and that most individuals diagnosed do not seek treatment, it is vital for more accessible modes of treatment delivery to be assessed for their efficacy.

Despite improving the accessibility of treatment delivery, an important point that may warrant further analysis is *who* is administering the treatment. In this analysis, the smallest treatment effect was demonstrated in the study by ter Huurne et al [30]. Importantly, this was the only study that did not have clinical psychologists guiding the treatment. It may be that the specialized experience of a clinical psychologist has a stronger impact on the treatment compared with less experienced medical health professionals. However, research has determined that the efficacy of treatment, including CBT, for other mental health disorders is similar between clinical psychologists and appropriately trained medical health professionals [58-61]. It

XSL•FO
RenderX

is important for future studies to determine whether these results can be generalized to patients with BED.

Since January 2019, 3 RCT study protocols have been published that outline the use of eHealth interventions for the treatment of BED [15,62,63]. All 3 propose using an internet-based form of CBT treatment and will use the DSM-5 criteria to diagnose BED. This is in contrast to this meta-analysis, in which of the 3 published studies, 2 used the DSM-IV criteria. One of the added benefits of using the DSM-5 criteria is that it may also increase participant recruitment for these studies. It has already been demonstrated that many health care providers and psychiatrists have difficulty identifying BED symptoms, leading to a greater need to improve the knowledge of the diagnostic criteria for BED [64]. With the inclusion of BED as a stand-alone diagnosis and the resulting increase in awareness of the disorder, participant recruitment for these studies may increase, providing a more accurate depiction of the efficacy of these eHealth treatments. This was also a phenomenon observed in the meta-analysis; when observing the effects of individual studies, the study with the largest sample size [10] was also the one that used the DSM-5 criteria. As a result, this study also found the largest treatment effect and narrowest CIs for all outcomes when compared with the other studies used in the analysis.

Limitations

One of the limitations of this study is its level of generalizability. The majority of participants were middle-aged women who were overweight or obese. Therefore, how well these findings can be applied to other age groups and male patients is not clear. It is important to note, however, that one of the reasons why the participant pool consists primarily of overweight and obese women may be that the disorder has a higher prevalence in females, and those diagnosed are 3-6 times more likely to be obese [65]. Furthermore, the higher prevalence of the disorder in women may enhance the efficacy and use of these eHealth treatments, as women are more likely to use the internet for medical and health-related information [66]. Another limitation of the study was that despite similarities in the study designs, heterogeneity was still quite high. The heterogeneity of the outcomes BMI, shape, and weight concern were in the moderate range; however, for the outcomes OBE and EDE-Q total, their values indicated substantial heterogeneity (77% for both). There are several reasons why this might have occurred. First, there was considerable variability in the criteria used to diagnose BED among the studies included in the analysis. In total, 2 of the studies used the DSM-IV criteria (1 relying on participant self-report) and 1 used the DSM-5 criteria. Compared with the DSM-IV, the use of the DSM-5 yields higher prevalence rates of BED and more accurate criteria for diagnosis [67-69]. In the analysis, the only study using the DSM-5 criteria not only had a higher sample size but also the largest effect size. Second, although the CBT-I and GSH-I treatments in the study were all based on CBT principles, the treatment protocols and duration were quite different. Finally, because of the limited number of studies and the limited participant pool, the results must be interpreted with caution.

Conclusions

This study provides preliminary evidence that eHealth treatments, and more specifically internet-based guided CBT treatments, are appropriate treatment avenues for BED. However, because of the limited number of published RCTs in this field, it is important for the current evidence base to become more complete, so that more conclusive results can be extracted. As more findings are published in this area, future studies not only need to analyze the efficacy of eHealth treatments but to further hone in on the effectiveness of in-person versus eHealth treatments, to investigate the differences in efficacy among the different types of eHealth treatments, to evaluate which elements of the treatment result in unchanged BMI, and to determine the characteristics of patients with BED that make them more suitable candidates for this alternative form of treatment.

Acknowledgments

The study was conceived by EM, who conducted the literature search and quantitative analysis, wrote the first draft of the manuscript, and made subsequent revisions. CD and MR assisted with the revision of the manuscript and addition of literature references, and MR assisted with quantitative analysis. All authors have approved the submission of the manuscript.

Conflicts of Interest

EM is currently developing a mobile app that tracks psychological well-being and screens for pathological overeating in individuals who are looking to lose weight. This app is in the development phase. Although it is not a direct eHealth intervention, its tracking functionality may influence the efficacy of the treatment that the individual is undergoing. CD and MR have no conflicts of interest to declare.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 66 KB - jmir_v23i7e17874_app1.doc]

Multimedia Appendix 2
Forest plot of outcomes.
[DOCX File , 6567 KB - jmir_v23i7e17874_app2.docx]

XSL•FO
RenderX

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington: American Psychiatric Association; 2013.
2. Hudson JI, Hiripi E, Pope HG, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007 Feb 01;61(3):348-358 [FREE Full text] [doi: 10.1016/j.biopsych.2006.03.040] [Medline: 16815322]
3. Erskine HE, Whiteford HA. Epidemiology of binge eating disorder. Curr Opin Psychiatry 2018 Nov;31(6):462-470. [doi: 10.1097/YCO.0000000000000449] [Medline: 30113324]
4. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. Lancet 2011 Aug 27;378(9793):815-825. [doi: 10.1016/S0140-6736(11)60814-3] [Medline: 21872750]
5. Amianto F, Lavagnino L, Abbate-Daga G, Fassino S. The forgotten psychosocial dimension of the obesity epidemic. Lancet 2011 Nov 19;378(9805):e8. [doi: 10.1016/S0140-6736(11)61778-9] [Medline: 22098857]
6. Amianto F, Ottone L, Abbate Daga G, Fassino S. Binge-eating disorder diagnosis and treatment: a recap in front of DSM-5. BMC Psychiatry 2015 Apr 03;15:70 [FREE Full text] [doi: 10.1186/s12888-015-0445-6] [Medline: 25885566]
7. Watson HJ, Jangmo A, Smith T, Thornton LM, von Hausswolff-Juhlin Y, Madhoo M, et al. A register-based case-control study of health care utilization and costs in binge-eating disorder. J Psychosom Res 2018 May;108:47-53 [FREE Full text] [doi: 10.1016/j.jpsychores.2018.02.011] [Medline: 29602325]
8. Ágh T, Kovács G, Supina D, Pawaskar M, Herman BK, Vokó Z, et al. A systematic review of the health-related quality of life and economic burdens of anorexia nervosa, bulimia nervosa, and binge eating disorder. Eat Weight Disord 2016 Sep;21(3):353-364 [FREE Full text] [doi: 10.1007/s40519-016-0264-x] [Medline: 26942768]
9. Ricca V, Castellini G, Mannucci E, Lo Sauro C, Ravaldi C, Rotella CM, et al. Comparison of individual and group cognitive behavioral therapy for binge eating disorder. A randomized, three-year follow-up study. Appetite 2010 Dec;55(3):656-665. [doi: 10.1016/j.appet.2010.09.019] [Medline: 20870000]
10. Wagner B, Nagl M, Dölemeyer R, Klinitzke G, Steinig J, Hilbert A, et al. Randomized controlled trial of an internet-based cognitive-behavioral treatment program for binge-eating disorder. Behav Ther 2016 Jul;47(4):500-514. [doi: 10.1016/j.beth.2016.01.006] [Medline: 27423166]
11. Wilson GT. Treatment of binge eating disorder. Psychiatr Clin North Am 2011 Dec;34(4):773-783. [doi: 10.1016/j.psc.2011.08.011] [Medline: 22098803]
12. McElroy SL, Guerdjikova AI, Mori N, Munoz MR, Keck PE. Overview of the treatment of binge eating disorder. CNS Spectr 2015 Dec;20(6):546-556. [doi: 10.1017/S1092852915000759] [Medline: 26594849]
13. Coffino JA, Udo T, Grilo CM. Rates of help-seeking in US adults with lifetime DSM-5 eating disorders: prevalence across diagnoses and differences by sex and ethnicity/race. Mayo Clin Proc 2019 Aug;94(8):1415-1426 [FREE Full text] [doi: 10.1016/j.mayocp.2019.02.030] [Medline: 31324401]
14. Regan P, Cachelin FM, Minnick AM. Initial treatment seeking from professional health care providers for eating disorders: a review and synthesis of potential barriers to and facilitators of "first contact". Int J Eat Disord 2017 Mar;50(3):190-209. [doi: 10.1002/eat.22683] [Medline: 28134980]
15. Munsch S, Wyssen A, Vanhulst P, Lalanne D, Steinemann ST, Tuch A. Binge-eating disorder treatment goes online - feasibility, usability, and treatment outcome of an Internet-based treatment for binge-eating disorder: study protocol for a three-arm randomized controlled trial including an immediate treatment, a waitlist, and a placebo control group. Trials 2019 Feb 13;20(1):128 [FREE Full text] [doi: 10.1186/s13063-019-3192-z] [Medline: 30760299]
16. Hamatani S, Numata N, Matsumoto K, Sutoh C, Ibuki H, Oshiro K, et al. Internet-based cognitive behavioral therapy via videoconference for patients with bulimia nervosa and binge-eating disorder: pilot prospective single-arm feasibility trial. JMIR Form Res 2019 Oct 23;3(4):e15738 [FREE Full text] [doi: 10.2196/15738] [Medline: 31647472]
17. Yim SH, Schmidt U. Experiences of computer-based and conventional self-help interventions for eating disorders: a systematic review and meta-synthesis of qualitative research. Int J Eat Disord 2019 Oct;52(10):1108-1124. [doi: 10.1002/eat.23142] [Medline: 31343088]
18. Yim SH, Bailey E, Gordon G, Grant N, Musiat P, Schmidt U. Exploring participants' experiences of a web-based program for bulimia and binge eating disorder: qualitative study. J Med Internet Res 2020 Sep 23;22(9):e17880 [FREE Full text] [doi: 10.2196/17880] [Medline: 32965235]
19. Moghimi E, Davis CA, Rotondi M. eHealth treatments for compulsive overeating: a narrative review. Curr Addict Rep 2020 Sep 01;7(3):395-404. [doi: 10.1007/s40429-020-00312-0]
20. Eng TR. Population health technologies: emerging innovations for the health of the public. Am J Prev Med 2004 Apr;26(3):237-242 [FREE Full text] [doi: 10.1016/j.amepre.2003.12.004] [Medline: 15026105]
21. König H, Bleibler F, Friederich H, Herpertz S, Lam T, Mayr A, et al. Economic evaluation of cognitive behavioral therapy and internet-based guided self-help for binge-eating disorder. Int J Eat Disord 2018 Feb;51(2):155-164. [doi: 10.1002/eat.22822] [Medline: 29345848]

XSL•FO
RenderX

22. Watson HJ, McLagan N, Zerwas SC, Crosby RD, Levine MD, Runfola CD, et al. Cost-effectiveness of internet-based cognitive-behavioral treatment for bulimia nervosa: results of a randomized controlled trial. J Clin Psychiatry 2018;79(1):11314 [FREE Full text] [doi: 10.4088/JCP.16m11314] [Medline: 29228517]
23. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided internet-based CBT for common mental disorders. J Contemp Psychother 2013 May 21;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
24. Aardoom JJ, Dingemans AE, Spinhoven P, Van Furth EF. Treating eating disorders over the internet: a systematic review and future research directions. Int J Eat Disord 2013 Sep;46(6):539-552. [doi: 10.1002/eat.22135] [Medline: 23674367]
25. Sadeh-Sharvit S, Kim JP, Darcy AM, Neri E, Vierhile M, Robinson A, et al. Subgrouping the users of a specialized app for eating disorders. Eat Disord 2018;26(4):361-372. [doi: 10.1080/10640266.2018.1440043] [Medline: 29452025]
26. Shingleton RM, Richards LK, Thompson-Brenner H. Using technology within the treatment of eating disorders: a clinical practice review. Psychotherapy (Chic) 2013 Dec;50(4):576-582 [FREE Full text] [doi: 10.1037/a0031815] [Medline: 23527906]
27. Loucas CE, Fairburn CG, Whittington C, Pennant ME, Stockton S, Kendall T. E-therapy in the treatment and prevention of eating disorders: a systematic review and meta-analysis. Behav Res Ther 2014 Dec;63:122-131 [FREE Full text] [doi: 10.1016/j.brat.2014.09.011] [Medline: 25461787]
28. Pittock A, Hodges L, Lawrie SM. The effectiveness of internet-delivered cognitive behavioural therapy for those with bulimic symptoms: a systematic review : a review of iCBT treatment for bulimic symptoms. BMC Res Notes 2018 Oct 22;11(1):748 [FREE Full text] [doi: 10.1186/s13104-018-3843-2] [Medline: 30348226]
29. Fairburn C. Overcoming Binge Eating. New York, United States: The Guilford Press; 1995:1-247.
30. ter Huurne ED, de Haan HA, Postel MG, van der Palen J, VanDerNagel JE, DeJong CA. Web-based cognitive behavioral therapy for female patients with eating disorders: randomized controlled trial. J Med Internet Res 2015 Jun 18;17(6):e152 [FREE Full text] [doi: 10.2196/jmir.3946] [Medline: 26088580]
31. Carrard I, Crépin C, Rouget P, Lam T, Golay A, Van der Linden M. Randomised controlled trial of a guided self-help treatment on the internet for binge eating disorder. Behav Res Ther 2011 Aug;49(8):482-491. [doi: 10.1016/j.brat.2011.05.004] [Medline: 21641580]
32. Carrard I, Crépin C, Rouget P, Lam T, Van der Linden M, Golay A. Acceptance and efficacy of a guided internet self-help treatment program for obese patients with binge eating disorder. Clin Pract Epidemiol Ment Health 2011;7:8-18 [FREE Full text] [doi: 10.2174/1745017901107010008] [Medline: 21552482]
33. de Zwaan M, Herpertz S, Zipfel S, Svaldi J, Friederich H, Schmidt F, et al. Effect of internet-based guided self-help vs individual face-to-face treatment on full or subsyndromal binge eating disorder in overweight or obese patients: the INTERBED randomized clinical trial. JAMA Psychiatry 2017 Oct 01;74(10):987-995. [doi: 10.1001/jamapsychiatry.2017.2150] [Medline: 28768334]
34. Shapiro JR, Reba-Harrelson L, Dymek-Valentine M, Woolson SL, Hamer RM, Bulik CM. Feasibility and acceptability of CD-ROM-based cognitive-behavioural treatment for binge-eating disorder. Eur Eat Disord Rev 2007 May;15(3):175-184. [doi: 10.1002/erv.787] [Medline: 17676687]
35. Juarascio AS, Goldstein SP, Manasse SM, Forman EM, Butryn ML. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
36. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington: American Psychiatric Association; 1994.
38. Striegel-Moore RH, Dohm FA, Solomon EE, Fairburn CG, Pike KM, Wilfley DE. Subthreshold binge eating disorder. Int J Eat Disord 2000 Apr;27(3):270-278. [doi: 10.1002/(sici)1098-108x(200004)27:3<270::aid-eat3>3.0.co;2-1] [Medline: 10694712]
39. Fairburn C, Cooper Z, O'Connor M. Eating disorder examination. In: Fairburn CG, editor. Cognitive Behavior Therapy and Eating Disorders. New York, United States: The Guilford Press; 2008:265-308.
40. Lewer M, Kosfelder J, Michalak J, Schroeder D, Nasrawi N, Vocks S. Effects of a cognitive-behavioral exposure-based body image therapy for overweight females with binge eating disorder: a pilot study. J Eat Disord 2017;5:43 [FREE Full text] [doi: 10.1186/s40337-017-0174-y] [Medline: 29296280]
41. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
42. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
43. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

XSL•FO
RenderX

44. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016 Mar;7(1):55-79 [FREE Full text] [doi: 10.1002/jrsm.1164] [Medline: 26332144]
45. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
46. Hilbert A, Petroff D, Herpertz S, Pietrowsky R, Tuschen-Caffier B, Vocks S, et al. Meta-analysis on the long-term effectiveness of psychological and medical treatments for binge-eating disorder. Int J Eat Disord 2020 Sep;53(9):1353-1376. [doi: 10.1002/eat.23297] [Medline: 32583527]
47. Hilbert A, Petroff D, Herpertz S, Pietrowsky R, Tuschen-Caffier B, Vocks S, et al. Meta-analysis of the efficacy of psychological and medical treatments for binge-eating disorder. J Consult Clin Psychol 2019 Jan;87(1):91-105. [doi: 10.1037/ccp0000358] [Medline: 30570304]
48. Traviss-Turner GD, Philpot U, Wilton J, Green K, Heywood-Everett S, Hill AJ. Guided self-help to manage binge eating in a dietetic-led community weight management service. Clin Obes 2018 Aug;8(4):250-257. [doi: 10.1111/cob.12259] [Medline: 29900680]
49. Peat CM, Berkman ND, Lohr KN, Brownley KA, Bann CM, Cullen K, et al. Comparative effectiveness of treatments for binge-eating disorder: systematic review and network meta-analysis. Eur Eat Disord Rev 2017 Sep;25(5):317-328. [doi: 10.1002/erv.2517] [Medline: 28467032]
50. Ghaderi A, Odeberg J, Gustafsson S, Råstam M, Brolund A, Pettersson A, et al. Psychological, pharmacological, and combined treatments for binge eating disorder: a systematic review and meta-analysis. PeerJ 2018;6:e5113 [FREE Full text] [doi: 10.7717/peerj.5113] [Medline: 29942715]
51. Cooper Z, Calugi S, Grave R. Controlling binge eating and weight: a treatment for binge eating disorder worth researching? Eat Weight Disord 2020 Aug;25(4):1105-1109. [doi: 10.1007/s40519-019-00734-4] [Medline: 31214963]
52. Quilty LC, Allen TA, Davis C, Knyahnytska Y, Kaplan AS. A randomized comparison of long acting methylphenidate and cognitive behavioral therapy in the treatment of binge eating disorder. Psychiatry Res 2019 Mar;273:467-474. [doi: 10.1016/j.psychres.2019.01.066] [Medline: 30684794]
53. Brownley KA, Berkman ND, Sedway JA, Lohr KN, Bulik CM. Binge eating disorder treatment: a systematic review of randomized controlled trials. Int J Eat Disord 2007 May;40(4):337-348. [doi: 10.1002/eat.20370] [Medline: 17370289]
54. Masheb RM, Dorflinger LM, Rolls BJ, Mitchell DC, Grilo CM. Binge abstinence is associated with reduced energy intake after treatment in patients with binge eating disorder and obesity. Obesity (Silver Spring) 2016 Dec;24(12):2491-2496 [FREE Full text] [doi: 10.1002/oby.21664] [Medline: 27797154]
55. Chao AM, Wadden TA, Walsh OA, Gruber KA, Alamuddin N, Berkowitz RI, et al. Perceptions of a large amount of food based on binge-eating disorder diagnosis. Int J Eat Disord 2019 Jul;52(7):801-808 [FREE Full text] [doi: 10.1002/eat.23076] [Medline: 30927476]
56. Yanovski SZ, Leet M, Yanovski JA, Flood M, Gold PW, Kissileff HR, et al. Food selection and intake of obese women with binge-eating disorder. Am J Clin Nutr 1992 Dec;56(6):975-980. [doi: 10.1093/ajcn/56.6.975] [Medline: 1442665]
57. Walsh BT, Boudreau G. Laboratory studies of binge eating disorder. Int J Eat Disord 2003;34 Suppl:30-38. [doi: 10.1002/eat.10203] [Medline: 12900984]
58. van Boeijen CA, van Oppen P, van Balkom AJ, Visser S, Kempe PT, Blankenstein N, et al. Treatment of anxiety disorders in primary care practice: a randomised controlled trial. Br J Gen Pract 2005 Oct;55(519):763-769 [FREE Full text] [Medline: 16212851]
59. Shandley K, Austin DW, Klein B, Pier C, Schattner P, Pierce D, et al. Therapist-assisted, internet-based treatment for panic disorder: can general practitioners achieve comparable patient outcomes to psychologists? J Med Internet Res 2008;10(2):e14 [FREE Full text] [doi: 10.2196/jmir.1033] [Medline: 18487138]
60. Høifødt RS, Strøm C, Kolstrup N, Eisemann M, Waterloo K. Effectiveness of cognitive behavioural therapy in primary health care: a review. Fam Pract 2011 Oct;28(5):489-504 [FREE Full text] [doi: 10.1093/fampra/cmr017] [Medline: 21555339]
61. Hiltunen AJ, Kocys E, Perrin-Wallqvist R. Effectiveness of cognitive behavioral therapy: an evaluation of therapies provided by trainees at a university psychotherapy training center. Psych J 2013 Aug;2(2):101-112 [FREE Full text] [doi: 10.1002/pchj.23] [Medline: 24436779]
62. Anastasiadou D, Lupiañez-Villanueva F, Faulí C, Cunillera J, Serrano-Troncoso E. Cost-effectiveness of the mobile application TCApp combined with face-to-face CBT treatment compared to face-to-face CBT treatment alone for patients with an eating disorder: study protocol of a multi-centre randomised controlled trial. BMC Psychiatry 2018 May 02;18(1):118 [FREE Full text] [doi: 10.1186/s12888-018-1664-4] [Medline: 29716580]
63. Vollert B, Beintner I, Musiat P, Gordon G, Görlich D, Nacke B, et al. Using internet-based self-help to bridge waiting time for face-to-face outpatient treatment for bulimia nervosa, binge eating disorder and related disorders: study protocol of a randomized controlled trial. Internet Interv 2019 Apr;16:26-34 [FREE Full text] [doi: 10.1016/j.invent.2018.02.010] [Medline: 30775262]

XSL•FO
RenderX

64. Chao AM, Rajagopalan AV, Tronieri JS, Walsh O, Wadden TA. Identification of binge eating disorder criteria: results of a national survey of healthcare providers. J Nurs Scholarsh 2019 Jul;51(4):399-407 [FREE Full text] [doi: 10.1111/jnu.12468] [Medline: 30821428]
65. Kessler RC, Berglund PA, Chiu WT, Deitz AC, Hudson JI, Shahly V, et al. The prevalence and correlates of binge eating disorder in the World Health Organization World Mental Health Surveys. Biol Psychiatry 2013 May 01;73(9):904-914 [FREE Full text] [doi: 10.1016/j.biopsych.2012.11.020] [Medline: 23290497]
66. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
67. Udo T, Grilo CM. Prevalence and correlates of DSM-5-defined eating disorders in a nationally representative sample of U.S. adults. Biol Psychiatry 2018 Sep 01;84(5):345-354 [FREE Full text] [doi: 10.1016/j.biopsych.2018.03.014] [Medline: 29859631]
68. Cossrow N, Pawaskar M, Witt EA, Ming EE, Victor TW, Herman BK, et al. Estimating the prevalence of binge eating disorder in a community sample from the United States: comparing DSM-IV-TR and DSM-5 criteria. J Clin Psychiatry 2016 Aug;77(8):968-974 [FREE Full text] [doi: 10.4088/JCP.15m10059] [Medline: 27232527]
69. Hudson JI, Coit CE, Lalonde JK, Pope HG. By how much will the proposed new DSM-5 criteria increase the prevalence of binge eating disorder? Int J Eat Disord 2012 Jan;45(1):139-141. [doi: 10.1002/eat.20890] [Medline: 22170026]

Abbreviations

BED: binge eating disorder
CBT: cognitive behavioral therapy
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edition
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ED: eating disorder
EDE-Q: Eating Disorder Examination Questionnaire
I-CBT: internet-based cognitive behavioral therapy
I-GSH: internet-based guided self-help
MeSH: Medical Subject Headings
OBE: objective binge episode
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
RCT: randomized controlled trial
SMD: standardized mean difference
WL: waitlist

Edited by R Kukafka; submitted 20.01.20; peer-reviewed by A Martinez-Millana, DM Liou, B Taylor; comments to author 06.09.20; revised version received 18.12.20; accepted 13.05.21; published 20.07.21.

Please cite as:
Moghimi E, Davis C, Rotondi M
The Efficacy of eHealth Interventions for the Treatment of Adults Diagnosed With Full or Subthreshold Binge Eating Disorder: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(7):e17874
URL: https://www.jmir.org/2021/7/e17874
doi:10.2196/17874
PMID:34283028

©Elnaz Moghimi, Caroline Davis, Michael Rotondi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy, Use, and Acceptability of a Web-Based Self-management Intervention Designed to Maximize Sexual Well-being in Men Living With Prostate Cancer: Single-Arm Experimental Study

Sean R O'Connor[1], PhD; Carrie Flannagan[2], PhD; Kader Parahoo[2], PhD; Mary Steele[3], PhD; Samantha Thompson[4], BSc; Suneil Jain[1,5], BAO, MB, MCRP, FRCP, PhD; Michael Kirby[6,7], MBBS, LRCP, MRCS, MRCP, FRCP; Nuala Brady[8], PhD; Roma Maguire[9], PhD; John Connaghan[9], MSc; Eilis M McCaughan[2], PhD

[1]Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom

[2]Institute of Nursing & Health Research, Ulster University, Newtownabbey, United Kingdom

[3]Centre for Clinical and Community Applications of Health Psychology, Faculty of Social and Human Sciences, University of Southampton, Southampton, United Kingdom

[4]Urology Department, Belfast City Hospital, Belfast, United Kingdom

[5]Clinical Oncology, Northern Ireland Cancer Centre, Belfast, United Kingdom

[6]Faculty of Health and Human Sciences, University of Hertfordshire, Hatfield, United Kingdom

[7]The Prostate Centre, London, United Kingdom

[8]Northern Health and Social Care Trust, Antrim, United Kingdom

[9]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, United Kingdom

Corresponding Author:
Sean R O'Connor, PhD
Centre for Public Health
Queen's University Belfast
Institute of Clinical Science, Royal Victoria Hospital
Belfast, BT12 6BA
United Kingdom
Phone: 44 28 9097 6350
Email: s.oconnor@qub.ac.uk

Abstract

Background: Sexual dysfunction is a frequent side effect associated with different prostate cancer treatment approaches. It can have a substantial impact on men and their partners and is associated with increased psychological morbidity. Despite this, sexual concerns are often not adequately addressed in routine practice. Evidence-based web-based interventions have the potential to provide ongoing information and sexual well-being support throughout all stages of care.

Objective: The aim of this study is to examine the efficacy of a web-based self-management intervention designed to maximize sexual well-being in men living with prostate cancer and explore user perspectives on usability and acceptability.

Methods: We used a single-arm study design, and participants were provided with access to the 5-step intervention for a period of 3 months. The intervention content was tailored based on responses to brief screening questions on treatment type, relationship status, and sexual orientation. Efficacy was assessed by using two-tailed, paired sample *t* tests for comparing the mean differences between pre- and postintervention measurements for exploring the participants' self-reported knowledge and understanding, sexual satisfaction, and comfort in discussing sexual issues. Usability and acceptability were determined based on the program use data and a postintervention survey for exploring perceived usefulness.

Results: A total of 109 participants were recruited for this study. Significant postintervention improvements at follow-up were observed in the total scores (out of 20) from the survey (mean 12.23/20 points, SD 2.46 vs mean 13.62/20, SD 2.31; t_{88}=9.570; *P*=.001) as well as in individual item scores on the extent to which the participants agreed that they had sufficient information to manage the impact that prostate cancer had on their sex life (mean 2.31/4 points, SD 0.86 vs mean 2.57/4, SD 0.85; t_{88}=3.660; *P*=.001) and had the potential to have a satisfying sex life following treatment (mean 2.38/4 points, SD 0.79 vs mean 3.17/4, SD 0.78; t_{88}=7.643; *P*=.001). The median number of intervention sessions was 3 (range 1-11), and intervention sessions had a median

XSL•FO
RenderX

duration of 22 minutes (range 8-77). Acceptable usability scores were reported, with the highest result observed for the question on the extent to which the intervention provided relevant information.

Conclusions: This study provides evidence on the efficacy of a tailored web-based intervention for maximizing sexual well-being in men living with prostate cancer. The results indicate that the intervention may improve one's self-perceived knowledge and understanding of how to manage sexual issues and increase self-efficacy or the belief that a satisfactory sex life could be achieved following treatment. The findings will be used to refine the intervention content before testing as part of a larger longitudinal study for examining its effectiveness.

(J Med Internet Res 2021;23(7):e21502) doi:10.2196/21502

KEYWORDS

prostate cancer; sexual well-being; digital interventions; self-management

Introduction

Background

Prostate cancer is the most common cancer among men, accounting for approximately 25% of all cases [1,2]. Although incidence rates are rising, partly because of improved screening and changes in the population age profile [3], 5- and 10-year survival rates continue to improve [4]. Consequently, an increasing number of men are living with significant long-term side effects associated with different treatment approaches [3]. Sexual challenges are the most frequently occurring sequelae [5,6]. Rates of sexual dysfunction having a moderate to severe impact on quality of life of 31%-64% of the men have been reported after radical prostatectomy and external beam radiotherapy [7,8]. In a recent large-scale survey, 81% of the men reported poor sexual function after treatment, with approximately 56% not being offered any intervention to help manage these concerns [9]. Changes in sexual function are subsequently regarded as a major issue that can result in increased psychological morbidity, including depression and relational dissatisfaction, and reductions in self-efficacy and overall quality of life [10]. Sexual well-being can be described as a complex and highly individualized issue that encapsulates all aspects of sexuality, including physical, emotional, mental, and social aspects [11,12]. Patients and their partners often have complex sexual health and well-being needs following diagnosis and treatment [13,14]. Effective evidence-based care and support are therefore required to help manage these needs. Care and support aimed at maximizing sexual well-being should not be restricted to purely biomedical approaches that focus on erectile dysfunction and physiologic penile rehabilitation [15]. These approaches do not address sexual well-being after prostate cancer diagnosis in a biopsychosocial context [16].

Although there is evidence examining relatively intensive couple-based counseling interventions delivered by health professionals [17,18], there is often limited access to such services. Current treatment guidelines [19,20] endorse the delivery of psychosexual care for patients living with prostate cancer with recommendations made for the minimal level of support that should be provided. This includes provision of individualized information tailored to the patients' needs and clear advice about the potential long-term side effects of treatment as well as ensuring ongoing access to specialist care, including erectile dysfunction clinics. Despite these recommendations, the information provided varies greatly and is not routinely available across services [9]. Patients and their partners frequently report that they do not receive adequate support to manage these concerns [21,22]. In a study of prostate cancer follow-up at urology and radiotherapy clinics, the sexual aspects of recovery were not discussed in 46% and 48%, respectively, of the observed consultations [23]. The participants' partners were present in approximately half of the consultations, but their involvement was minimal, and they did not seem to influence whether any discussion of sexual concerns took place [23].

Discussing sexual health concerns in routine practice can be challenging, and there are a number of barriers to engaging in these conversations [24,25]. Health care providers often feel unequipped to deal with sexual health issues and report a lack of resources to offer patients and their partners if they do identify a problem [26]. Patients may not spontaneously report sexual health issues and prefer that health professionals initiate the discussion [27]. These assumptions may be compounded when health professionals work with patients from minority groups such as men who have sex with men. For example, many gay men report that health professionals often fail to ask about sexual orientation during the initial consultations and assume that they are heterosexual [28].

Web-based interventions provide access to ongoing, easily accessible, and adaptable information and support to users at all stages of care [29]. There is evidence that the tailoring or the personalization of web-based information and support interventions is more effective and results in increased user engagement when compared with standardized information [30]. In addition, tailored self-management interventions are more capable of altering determinants of individual beliefs and behaviors [31]. However, some barriers exist that can limit engagement with web-based resources, including a lack of time and usability issues [32]. Despite this, web-based interventions that specify and acknowledge the impact of treatment on the sexual well-being of both men and their partners and provide appropriate support have the potential to improve patient-important outcomes, including sexual well-being satisfaction and quality of life. Such interventions, which are aimed at supporting men and their partners to cope with changes in sexual health and well-being after prostate cancer treatment, require further investigation.

This paper presents an evaluation of a web-based self-management intervention designed to maximize sexual

XSL•FO
RenderX

well-being in men living with prostate cancer. The program provides tailored information and support based on the user's treatment type, relationship status, and sexual orientation. This aligns with existing guidelines that advocate tailored psychosexual interventions [19,20]. It is also in line with recommendations that emphasize early support, consisting of educational approaches and interventions to manage sexual side effects of treatment and minimize the impact of changes to sexual function on the men and their partners [33].

Objectives

Recent frameworks for developing and evaluating complex health care interventions emphasize a requirement for greater focus on initial development because many fail to demonstrate effectiveness in real-world contexts [34]. Before conducting larger studies exploring intervention effectiveness, this study was conducted to examine if the intervention had any effect on patient-important outcomes and to explore its acceptance to users. Therefore, the primary objective of this study is to examine the efficacy of the intervention in terms of its impact on participants' understanding of how to manage sexual concerns, comfort in discussing such issues with partners and health professionals, and overall satisfaction with their sex life. The secondary objective is to explore program use and user perspectives on usability and acceptability.

Methods

Study Design

A single-arm pilot study design with pre- and postintervention outcome assessments was used. Following enrollment, the participants were given access to the intervention for a 3-month period. Where appropriate, the design and conduct of the study followed the Consolidated Standards of Reporting Trials 2010 statement: extension to randomized pilot and feasibility trials [35].

Study Setting and Participants

The primary study recruitment methods were through health professionals signposting to the study website men who were attending routine prostate cancer appointments at 2 clinical sites (Northern Ireland Cancer Centre, Belfast City Hospital, Belfast, United Kingdom, and Ninewells Hospital, Dundee, United Kingdom) and through posters and leaflets placed in clinical areas within the same sites. In addition, a link to the program was included in the patient information section of a national prostate cancer charity website. A minimum sample of 81 participants was determined based on two-tailed, paired sample *t* tests, α=.05, and a medium estimated effect size of 0.03 [36]. Therefore, a planned sample size of 100 participants was selected to allow for potential loss of data at follow-up. Following web-based registration on the site, potential participants were required to complete a screening questionnaire before a baseline assessment. To meet the study inclusion criteria, participants were required to be adult males (aged 18 years or older); diagnosed with prostate cancer; and due to start, or be currently receiving, supportive care after radical prostatectomy, external beam radiotherapy, brachytherapy, or androgen deprivation therapy (either alone or in combination). The exclusion criteria were as follows: being on active surveillance or not being able to understand instructions written in English.

Study Procedures

Ethical approval for the study was provided by the Office for Research Ethics Committees Northern Ireland (reference number: 17/NI/014). Before completing the web-based screening questionnaire, the participants were provided with a study information sheet detailing the nature and purpose of the study. They were also given the opportunity to contact a member of the research team to ask any questions they might have about the study. All participants provided informed consent before participation. Subsequently, they completed the baseline assessment, which included demographic information and baseline outcomes, and provided responses to the 3 questions that were used to enable the intervention to provide tailored information and support based on the responses given (Textbox 1). The participants were given access to the program for 3 months. The only contact that the participants received during the intervention period was through automated emails sent to confirm successful enrollment and to remind them that they had 1 week left to use the intervention. After the 3-month intervention period ended, the participants received email reminders asking them to log in to the website and complete a follow-up assessment in which the baseline outcomes were repeated. They were also asked to complete a questionnaire on program usability and acceptance.

XSL•FO
RenderX

Textbox 1. Questions asked at baseline to allow tailored information to be provided by the intervention.

Tailoring questions

- What treatment have you had?
  - Surgery
  - Combined radiotherapy and hormone therapy
  - Radiotherapy
  - Hormone therapy
- Are your sexual partners usually male or female?
  - Female
  - Male
- Do you currently have a partner?
  - Yes
  - No

Intervention Development, Theory, and Description

Intervention Development

A systematic, iterative, and theory-based process modeled on the person-based approach was used to inform the development, design, and testing [37]. This method was primarily used to ensure that the development was in close collaboration with end users and to optimize intervention acceptability, feasibility, and engagement. This process included 2 phases: an intervention development and testing phase and an evaluation and follow-up phase. The draft intervention content was modeled on an existing sexual well-being intervention [38]. In the first phase, evidence reviews and a qualitative synthesis of data from semistructured interviews and focus group discussions with end users and field content experts were used to identify the core or essential elements of the intervention. Additional interviews with both types of participants were then used to review and revise the paper-based versions of the content. This was to ensure that it was relevant and meaningful to users. An initial prototype version of the intervention was subsequently built using LifeGuide software (University of Southampton) [39], which provides tools for developers to author, edit, deploy, and trial interventions. Further modifications were made based on usability testing and additional rounds of qualitative interviews. These steps were carried out before making further revisions and building the final version of the intervention that was used for this evaluation. In the second phase, evaluation of the intervention was conducted based on quantitative and qualitative data exploring preliminary efficacy, use, and acceptability data, which are presented in this paper.

Theoretical Underpinning

As the intervention was delivered in a web-based format, its theoretical underpinning was based on the unified theory of acceptance and use of technology, a widely used model of technology acceptance and use intention [40]. This model integrates a number of relevant technology acceptance and behavior change theories, including self-efficacy, the theory of reasoned action, technology acceptance theory, the theory of planned behavior, and social cognitive theory. Critical to the theory of acceptance and use of technology model are the concepts of perceived usefulness and ease of use. The central determinants of intention and use are performance and effort expectancy, social influences, and facilitating conditions, with factors such as age, gender, prior experience, and voluntariness to use assumed to be moderators of these effects [41].

Intervention Description

The final version of the intervention consisted of a 5-step program designed to maximize sexual well-being in men living with prostate cancer. The 5 steps were as follows: (1) sexual well-being and prostate cancer, (2) changes and coping with changes, (3) maintaining and improving your sex life, (4) exploring sexual pleasure, and (5) facing the future. In addition, a user toolkit containing a series of quick guides was included. Each step varied in length from 12 to approximately 40 webpages. Information was also layered using page tabs, meaning that although all participants were required to view core information, other information could be skipped or viewed at a later date. The intervention provided tailored information and support based on the user's treatment type, relationship status, and sexual orientation, with the program allowing different information to appear on screen based on the responses given to the brief tailoring questions that the participants completed during initial registration. The participants were encouraged to use the program with their partner, and specific tailored information was included for partners. This included, for example, information for female partners on women's health and couple communication activities as well as advice on talking to a partner's health care team. It was recommended that the users complete the steps in sequence over the 3-month intervention period. However, a key design feature of the intervention was that all steps were accessible from the start of the intervention period. It was also emphasized to the participants that the intervention was designed as a resource that they could return to at any time to revisit previously viewed sections or to view or complete unfinished steps. The

XSL•FO
RenderX

participants received a tick mark over each step, which could be seen each time they logged in. This was to indicate the steps that they had already completed.

Each step consisted of a series of webpages containing text-based information, infographics, videos highlighting patient experiences, and instructional videos delivered by health care professionals. Some steps included exercises, activities, and other resources for participants. These included a couple communication activity and a printed checklist for the participants to use when discussing sexual issues or concerns with their health care professional. The intervention content also included important behavior change components and techniques, including use of social support; information about health consequences; instruction on how to perform a behavior; demonstration or modeling of behavior; and use of prompts, reminders, and cues [42]. The key principles and characteristics of the final intervention version are listed in Textbox 2. A screenshot of the intervention home page is shown in Figure 1.

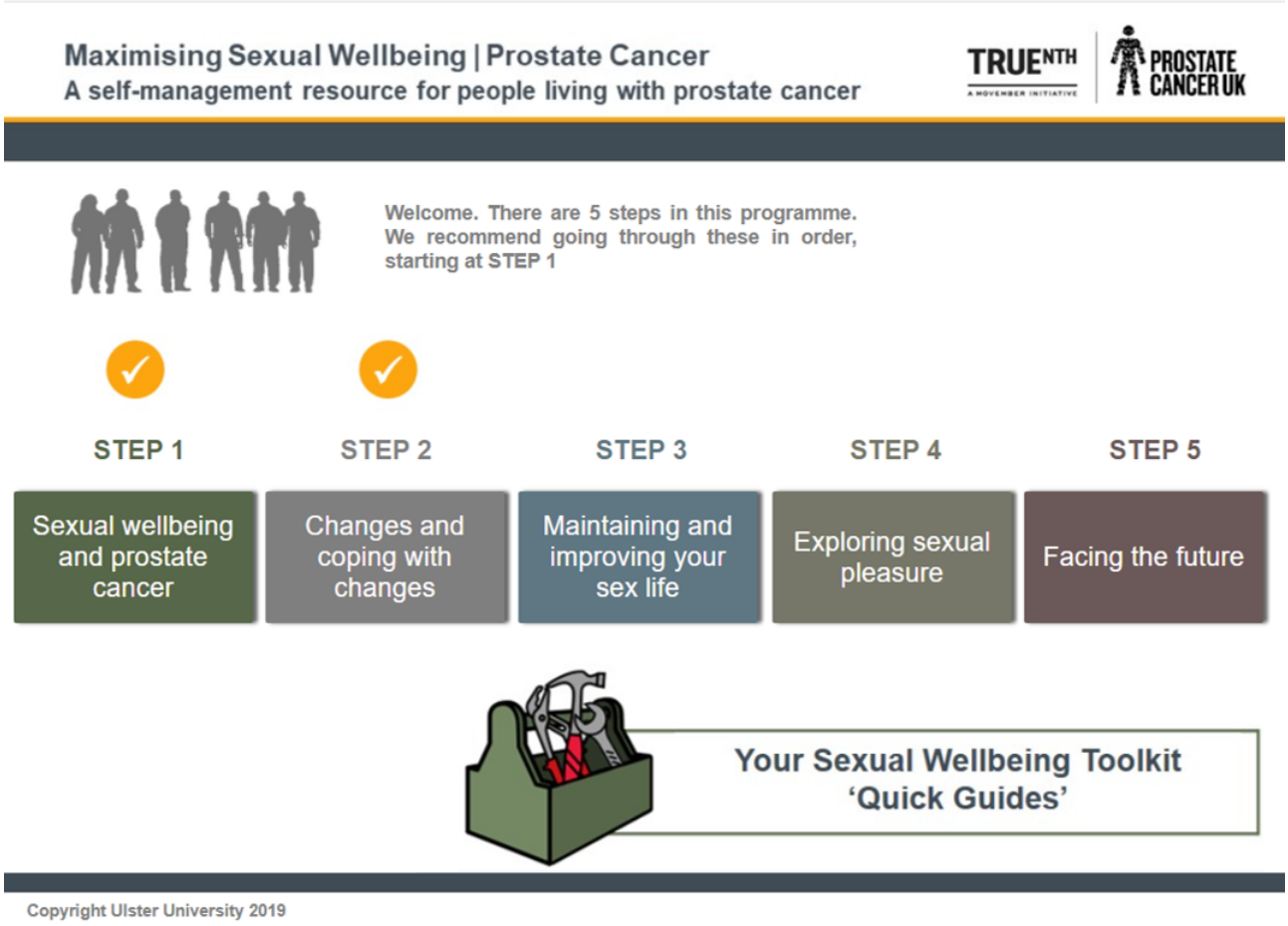
Textbox 2. Key principles and characteristics of the final intervention.

Key principles

- To normalize sexual concerns associated with prostate cancer and its treatment and address patient and partner expectations of potential sexual recovery
  - Provide case-based examples, including patient experience videos
  - Provide potential side effects information, including common methods of coping and managing individual side effects
- To acknowledge changes or potential loss of sexual function and promote resilience and effective coping strategies
  - Promote benefits of adapting to a new sexual normal and adopting new approaches and working as a couple
  - Provide instructional and demonstration videos presented by health professionals
- To provide personalized information and support based on needs, including treatment type, sexual orientation, and partner status
  - Provide layering of information and support based on needs (ensuring that the intervention can be used for brief periods but can also facilitate more in-depth or intensive support based on user needs)
- To promote increased sexual well-being conversations between partners and health professionals
  - Provide printable health professional communication aid
  - Provide printable couple communication exercise
  - Provide specific supportive information for partners to promote shared intervention use
- To provide usable, easily accessible, and relevant support available at all stages of care
  - Include printable exercises and activities to be used as prompts or reminders of key points
  - Provide information on appropriate support services based on needs
  - Use simple design interface with core information provided on main webpages and selected additional information available based on user preference

XSL•FO
RenderX

Figure 1. Intervention screenshot.



Outcomes

As the objectives of this study are to explore the efficacy of the intervention and examine use data and user perspectives on usability and acceptability, a 3-month pilot study was conducted. Efficacy was assessed using pre- and postintervention measurements of a self-reported web-based survey that included 1 question exploring knowledge and understanding, 2 questions on sexual satisfaction, and 2 questions on comfort in discussing sexual issues (with health care professionals and with a partner). The participants were asked to rate their level of agreement with the 5 different questions using a 4-point Likert scale anchored by *strongly disagree* and *strongly agree* at either end. The composite efficacy score out of 20 was calculated by combining the scores for all 5 questions.

Intervention use was determined by calculating the number of intervention sessions (log-ins) for each participant, the duration of each session, and the total time spent using the intervention over the 3-month evaluation phase. In addition, the participants were classified as *completers* or *noncompleters* based on whether they had completed at least 4 of the 5 intervention steps.

Usability and acceptability were determined based on different methods, including a brief web-based survey and free-text responses to 2 questions asked at the 3-month follow-up assessment. This survey was based on a modified and shortened version of the system usability scale [43]. Modifications were made to ensure that the questions were relevant to the assessment of intervention acceptability. This included the addition of questions on the look and design of the program and the relevance of the information provided. The participants were asked to rate their level of agreement with each of the 6 questions using a 4-point Likert scale anchored by *strongly disagree* and *strongly agree* at either end. A composite usability score out of 24 was calculated by combining the scores for all 6 questions. The participants were also asked whether they would recommend the intervention to others (yes, not sure, or no). Finally, the participants were asked to provide free-text responses to the following questions:

1. Did you gain anything from using the intervention?
2. Do you have any recommendations on how the intervention can be improved?

Data Analysis

Data were exported into SPSS version 25.0 (IBM Corporation), which was used to provide a descriptive analysis of demographic data, intervention use data, and usability ratings. To assess intervention efficacy, paired sample *t* tests were used to compare the mean pre- and postintervention efficacy measures for the composite and individual question scores. Data were tested for normality of distribution, and a Bonferroni-adjusted P=.007 (P value of .05 divided by the number of comparisons: n=6) was used to allow for multiple comparisons. Estimated effect sizes for pre-post intervention effects were also calculated using the following criteria: 0.00-0.19, insignificant; 0.20-0.49, small;

XSL•FO
RenderX

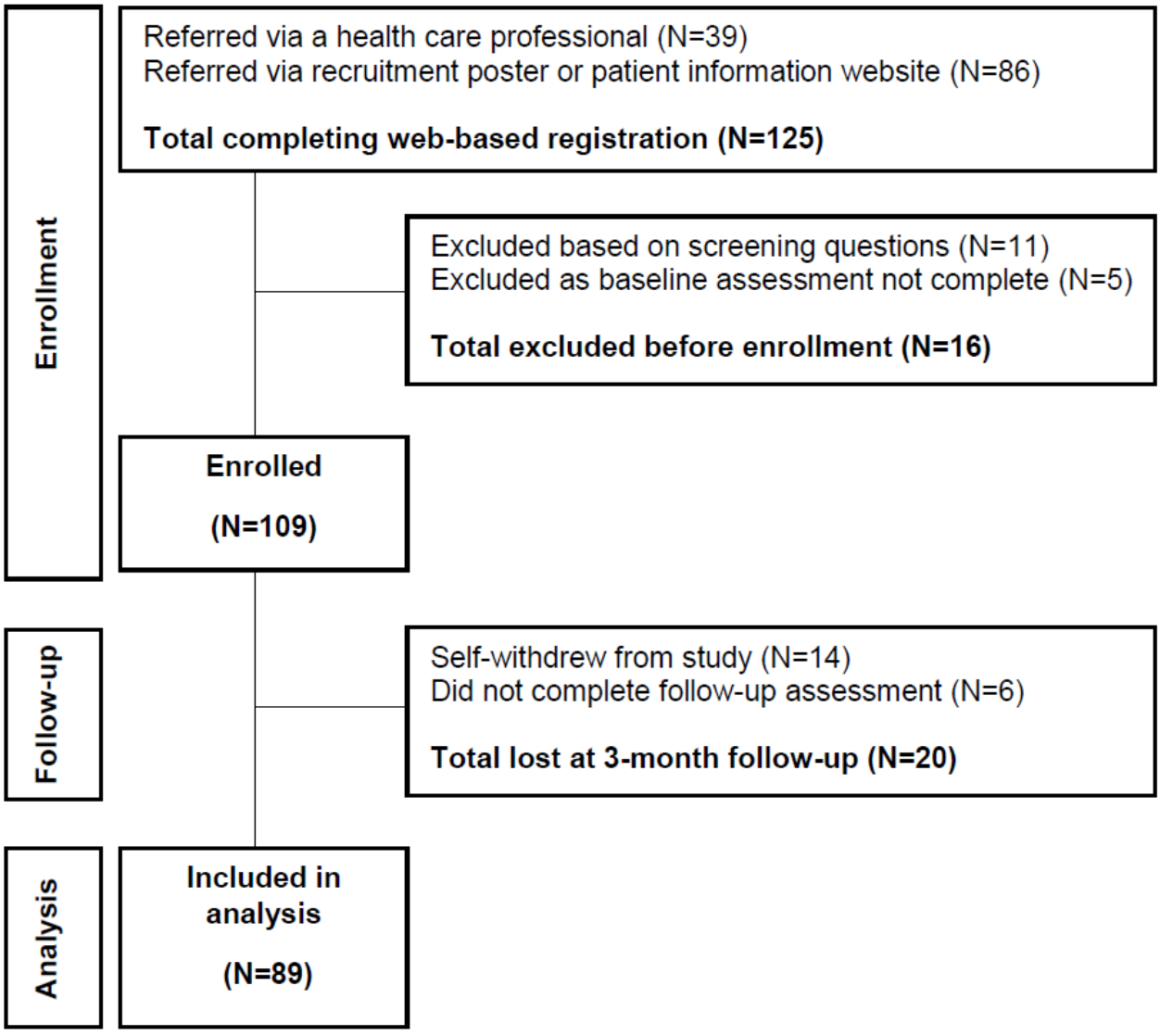
0.50-0.79, medium; and ≥0.80, large [44]. Independent sample *t* tests were then used to test for any significant differences between the composite and individual usability question scores between the participants classified as *completers* (those who accessed at least 4 of the 5 intervention steps) and those classified as *noncompleters*.

Results

Participant Flow and Retention

Participants were recruited for the study between February 2019 and July 2019. A total of 125 potential participants completed the initial web-based registration; however, 16 of these participants were not enrolled on the basis of their responses to the screening questions or because they did not complete the baseline assessment questions. Therefore, a total of 109 men were enrolled in the study, and they provided informed consent to participate (Figure 2). Of the 109 men, 20 (18.3%) were lost to follow-up at 3 months owing to self-withdrawal (defined as no intervention use or log-ins after initial registration) or noncompletion of follow-up assessment questionnaires; this resulted in data from 89 (81.7%) participants being included in the final analysis.

Figure 2. Participant flow diagram.



Participant Demographics

Most men (66/89, 74%) were aged between 50 and 69 years, and most of them (83/89, 93%) were from a White ethnic background. Surgical intervention alone (not in combination with any other form of treatment) was the single most common form of treatment received (53/89, 60%), with combined treatment (including radiotherapy and hormone therapy) being the second most common form of treatment (20/89, 22%). Full demographic details of the participants are shown in Table 1. There were no observable differences in demographics between the 89 men included in the analysis and the 20 men who were lost to follow-up.

XSL•FO
RenderX

Table 1. Demographic details of participants enrolled in the study (N=109).

Demographics	Included in analysis (n=89), n (%)	Withdrawals (n=20), n (%)
Age category (years)		
18-49	1 (1)	0 (0)
50-69	66 (74)	16 (80)
≥70	22 (25)	4 (20)
Ethnicity		
White	83 (93)	19 (95)
Asian or Asian British	1 (1)	0 (0)
Black, African, Caribbean, or Black British	4 (5)	1 (5)
Other	1 (1)	0 (0)
Previous sexual care or support received		
Yes	58 (65)	9 (45)
No	31 (35)	11 (55)
Timing of any previous sexual care or support received		
At diagnosis	16 (18)	5 (56)[a]
During treatment	10 (11)	4 (44)[a]
Both	32 (37)	0 (0)[a]
Type of prostate cancer treatment received		
Surgery only	53 (60)	17 (85)
Radiotherapy only	9 (10)	0 (0)
Hormone therapy only	7 (8)	0 (0)
Combined therapy	20 (22)	2 (10)
Treatment phase		
Pretreatment or on ongoing treatment	20 (23)	16 (80)
Less than 6 months of completing treatment	37 (42)	4 (20)
More than 6 months after completing treatment	32 (36)	0 (0)
In a relationship		
Yes	82 (91)	20 (100)
No	7 (6)	0 (0)
Usual partner gender		
Female	81 (91)	20 (100)
Male	8 (9)	0 (0)

[a]n=9.

Efficacy Data

The data are normally distributed. On the basis of the mean differences in pre- and postintervention (3 months) self-reported measures, a significant improvement was observed in total composite efficacy scores (t=9.570; P=.001), with a medium estimated effect size (Cohen d=0.577; Table 2). For individual survey items, significant improvements were seen in mean scores for (1) participants' understanding of how to manage the impact of treatment (t=3.660; P=.001) and (2) participants' perceptions of their ability to maintain a satisfying sex life despite cancer treatment (t=7.643; P=.001). No significant effects were found for participants' mean current level of sexual satisfaction or level of comfort when discussing sexual issues with a partner or a health professional.

XSL•FO
RenderX

Table 2. Mean differences in pre- and postintervention (3 months) self-reported efficacy measures with estimated effect sizes[a].

Individual statement	Baseline score, mean (SD)	Score at 3 months, mean (SD)	*t* test (*df*)	*P* value[b] (2-tailed)	Effect size (Cohen *d*)	Effect size interpretation[c]
I currently have a satisfying sex life	1.89 (0.92)	1.90 (0.94)	0.376 (88)	.71	0.01	Insignificant
I have a good understanding of how to manage the impact of prostate cancer treatment on my sex life	2.31 (0.86)	2.57 (0.85)	3.660 (88)	.001[d]	0.517	Medium
I can have a satisfying sex life despite prostate cancer treatment	2.38 (0.79)	3.17 (0.78)	7.643 (88)	.001[d]	1.001	Large
I am comfortable discussing sexual issues with a partner	3.08 (0.72)	3.12 (0.7)	0.851 (88)	.40	0.055	Insignificant
I am comfortable discussing sexual issues with a health professional	2.81 (0.74)	2.82 (0.71)	0.241 (88)	.81	0.014	Insignificant

[a]Individual statements were scored on a scale between 1 and 4 points based on the response to the following: "How much do you agree with each statement?" Responses were measured on a 4-point scale anchored by *strongly disagree* and *strongly agree* at either end. Total composite scores out of 20 were calculated by combining scores from each statement. The total composite scores are as follows: mean baseline score, 12.23 (SD 2.46); mean score at 3 months, 13.62 (SD 2.31); t_{88}=9.570; *P*=.001; effect size (Cohen *d*)=0.577; and effect size interpretation, medium.

[b]Bonferroni-adjusted *P* value for multiple comparisons (*P*=.007).

[c]Effect size interpretation: 0.00-0.19 (insignificant), 0.20-0.49 (small), 0.50-0.79 (medium), and ≥0.80 (large).

[d]Denotes a significant pre-post intervention effect.

Intervention Use Data

An analysis of program use during the 3-month intervention phase indicated that engagement with the intervention varied, suggesting that the participants used the intervention differently based on their individual needs and preferences. The participants completed a median of 3 sessions (range 1-11). The median session duration was 22 minutes (range 8-77), with an overall total use time of 78 minutes (range 18-284; Table 3). Of the 89 participants, 45 (51%) completed at least 4 of the 5 intervention steps and were subsequently classified as *completers*. Although the number of sessions and duration of each session reduced each month during the intervention period (Figure 3 and Figure 4), 85% (76/89) and 65% (58/89) of the participants were still using the intervention in the second and third months, respectively, of the intervention phase.

Table 3. Median and mean values for program use data.

Use measure	Value, median (IQR)	Value, mean (SD)
Number of sessions	3 (4)	3.8 (1.98)
Duration of each session (minutes)	22 (18)	36.4 (16.7)
Duration of total use time (minutes)	78 (80)	115.2 (43.5)

XSL•FO
RenderX

Figure 3. Box plot showing the number of sessions in each month over the duration of the intervention period.

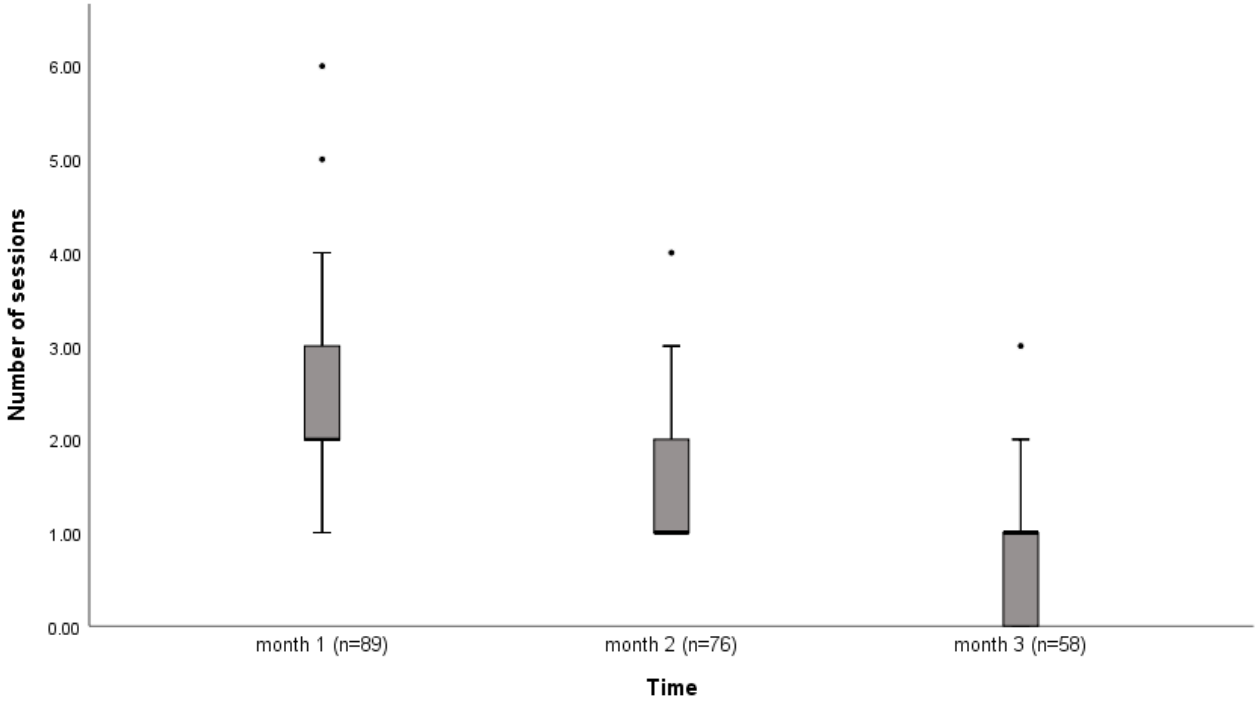
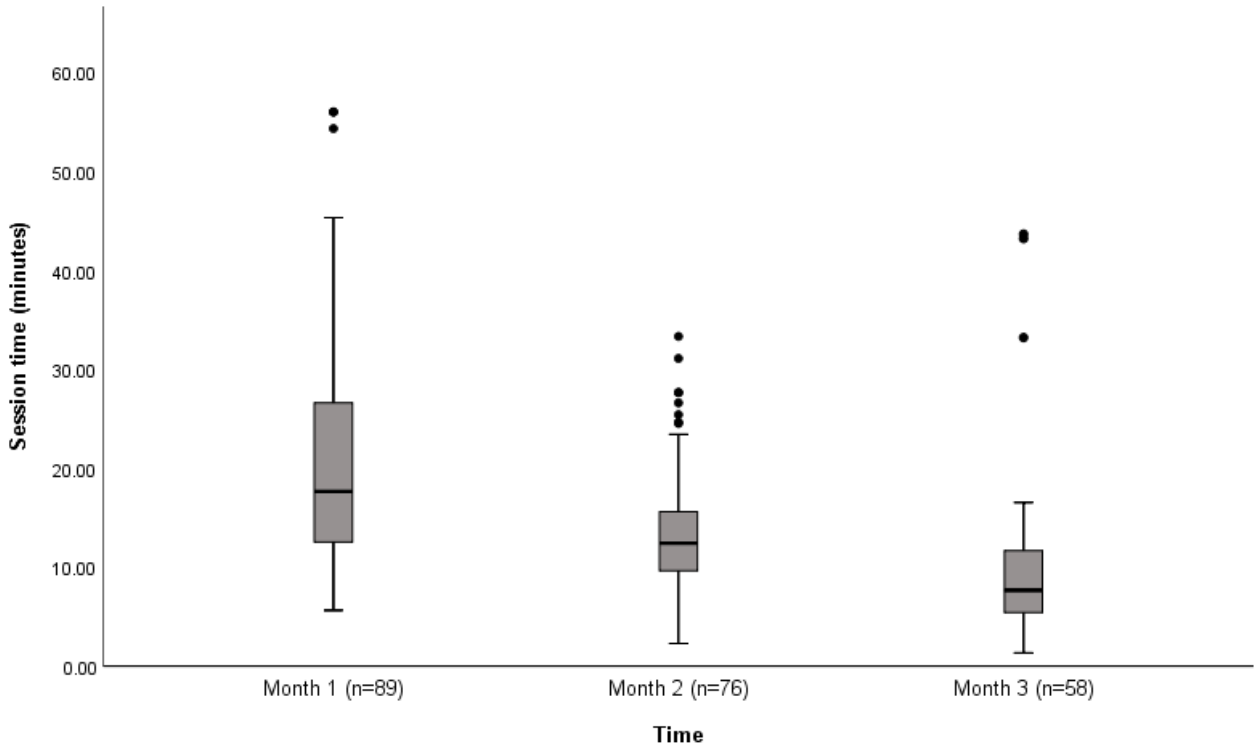


Figure 4. Box plot showing the duration of each session during the intervention period.



Usability and Acceptability Data

On the basis of the postintervention survey data, the overall usability scores were found to be acceptable (total composite score: 19.68/24, 82% agreement). The highest levels of agreement were observed for the questions on *trust in theprogramme* (3.36/4, 93% agreement) and *information included was useful to me* (3.77/4, 94% agreement). The lowest agreement scores were found for the questions *I liked the look of the programme* (2.87/4, 72% agreement) and *I found the programme easy to use* (3.03/4, 76% agreement). Two-tailed, independent sample *t* tests identified that there were no

XSL•FO
RenderX

significant differences in composite or individual question scores on the usability survey between the participants classified as *completers* and those classified as *noncompleters* (Table 4). Of the 89 participants, 70 (79%) agreed that they would recommend the intervention to others. The participants' responses to questions on what they gained from using the intervention and any recommendations on how it could be improved are summarized in Table 5.

Table 4. Mean and percentage agreement scores for usability survey data (N=89)[a].

Individual statement	Score, mean (SD)	Agreement (n=4), n (%)	Completers[b] (n=45), mean score (SD)	Noncompleters (n=44), mean score (SD)	*P* value (difference between completers and noncompleters)
I was satisfied with the program	3.16 (0.56)	3.16 (79)	3.21 (0.68)	3.11 (0.71)	.44
I found the program easy to use	3.03 (0.69)	3.03 (76)	3.15 (0.7)	2.91 (0.67)	.10
I was able to move through the program easily	3.49 (0.56)	3.49 (87)	3.57 (0.56)	3.42 (0.61)	.37
I liked the look of the program	2.87 (0.47)	2.87 (72)	2.89 (0.67)	2.85 (0.71)	.45
I felt I could trust the program	3.36 (0.55)	3.36 (93)	3.26 (0.54)	3.47 (0.65)	.34
The information included in the program was useful to me	3.77 (0.53)	3.77 (94)	3.84 (0.55)	3.71 (0.48)	.63

[a]Individual statements were scored on a scale between 1 and 4 points based on the response to the following: "How much do you agree with each statement?" Responses were measured on a 4-point scale anchored by *strongly disagree* and *strongly agree* at either end. Total composite scores out of 24 were calculated by combining scores from each statement. The total composite scores were as follows: mean score, 19.68 (SD 0.56); agreement (19.68/24, 82%); mean completer score, 19.92 (SD 0.61); mean noncompleter score, 19.47 (SD 0.63); and *P*=.24.

[b]Completers were defined as participants who completed at least 4 of the 5 steps of the web-based program.

XSL•FO
RenderX

Table 5. Summary of participant comments on what they gained from using the intervention and recommendations on how it could be improved (N=207).

Explanation	Comments, n (%)	Category
Views on what was gained from using intervention (n=141)		
The program provided information on issues not previously thought about	33 (23.4)	Information
The program provided useful warnings on possible effects of treatment	26 (18.4)	Information
The program helped to normalize sexual problems	22 (15.6)	Information and tone or language
The program provided new information not previously discussed with health professionals	20 (14.2)	Information
The program provided information that was relevant and useful to me as an individual	11 (7.8)	Information and personalization
The program helped provide ideas for different approaches to manage sexual problems	9 (6.4)	Information and confidence or self-efficacy
The program provided information that could be viewed and discussed with a partner	9 (6.4)	Information and communication
The program provided a positive tone and message, which was reassuring	5 (3.5)	Tone, language and confidence, or self-efficacy
The program helped to increase my confidence	4 (2.8)	Confidence or self-efficacy
The program provided a reminder of information that was previously discussed with health professionals	2 (1.4)	Information
Suggested improvements that could be made to the intervention (n=66)		
Make intervention available to patients before treatment starts	32 (48.5)	N/A[a]
Make intervention available as a mobile app	17 (25.7)	N/A
Include more support such as someone to contact for advice	14 (21.2)	N/A
Make intervention available in an offline or printed format	3 (4.5)	N/A

[a]N/A: not applicable.

Discussion

Principal Findings

The findings from this study provide evidence of the efficacy of a web-based intervention designed to maximize sexual well-being in men living with prostate cancer. An analysis of self-reported outcome data found that the intervention resulted in significant improvements at the 3-month follow-up in overall efficacy scores and participants' understanding of how to manage the impact of sexual concerns as well as their perceived ability to have a satisfying sex life despite prostate cancer treatment. The findings also indicated that the program had good overall usability and acceptability. Although the participants used the web-based self-management intervention in markedly different ways, they typically engaged well, taking part in multiple sessions during the intervention period. This is one of the first studies to evaluate the potential effectiveness and use of a tailored, sexual well-being support intervention for men living with prostate cancer and their partners, which is delivered using a web-based platform. A key strength of the intervention seems to be its flexibility with support that can be personalized based on the user's needs and delivered at any stage of care.

Intervention Efficacy

Self-reported measures at the 3-month follow-up were used to evaluate intervention efficacy, and they demonstrated significant overall improvements in comparison with the baseline scores. Improvements with a medium effect size were found in the extent to which the participants agreed that they had sufficient information to manage the impact of prostate cancer on their sex life. In addition, the extent to which the participants agreed that there was potential for them to have a satisfying sex life following treatment also improved significantly. The findings indicate that the intervention had a positive influence on the men's self-perceived knowledge and understanding of how to manage sexual issues, but, more importantly, it also seemed to contribute toward a substantial increase in self-efficacy or a belief that a satisfactory sex life could be achieved following treatment. This indicates a potentially important prerequisite for maintaining behavioral change. Higher coping self-efficacy can result in more effective responses to behavioral barriers or setbacks, with individuals more able to apply behavior change maintenance strategies such as action planning [45,46].

An examination of the individual survey item scores suggested that the intervention had no effects on the participants' current level of satisfaction with their sex life or on the level of comfort in discussing sexual issues with a partner or with a health care professional. The extent to which the participants agreed that they were happy with their current level of satisfaction was low

XSL•FO
RenderX

at baseline, and although this may have been a factor for motivating potential participants to take part in the study, the intervention did not lead to improvements in this measure. This may have been due to the comparatively short timescale of the evaluation phase. Changes in sexual function after treatment are dependent on treatment type [9], with many effects having a long-term or persistent impact. Coping with these changes and adapting new practices as part of an individual's sex life can take time. It may be necessary to use longitudinal studies to evaluate interventions aimed at improving satisfaction with current sex life, which is a complex, multifactorial concept that is closely related to the overall quality of life [47] and potentially mediating factors such as relationship status and expectations of recovery. The extent to which the participants agreed that they were comfortable when discussing sexual issues also did not change, but these scores were relatively high at baseline. This supports the findings from studies that have found that the level of comfort in men with prostate cancer is not a significant barrier to discussing sexual issues [21,48].

Use, Usability, and Acceptability

Overall, the findings indicated that although the participants accessed the intervention a median of 3 times, the patterns of use seemed to differ among the participants. For example, engagement varied with some using the intervention more frequently over a number of shorter sessions throughout the intervention phase and others using it a limited number of times but with longer intervention sessions. This is reflected in the wide range of session numbers and session durations of between 1 and 11 sessions and 8 and 77 minutes, respectively. Although use reduced over the intervention period, approximately 65% (58/89) of the participants still showed engagement with the program in the final month. Previous evidence has demonstrated levels of engagement with web-based interventions that are comparable with face-to-face delivery methods [49], and it has been suggested that web-based resources are viewed as an acceptable and widely used source of information on sexual concerns [50]. Various needs of web-based interventions have been identified, including improving couple communication and providing information on sexual side effects, rehabilitation approaches, and realistic expectations of recovery [51]. The reasons for the variation in user engagement in this study may be related to a number of factors, including the fact that the intervention was intended to be used differently based on users' individual needs and preferences. There is also evidence of an association between perceived usability or ease of use and engagement [52]. Usability and design issues could be additional reasons that may account for the different user engagement in this study. Although no significant differences were found among the users who completed at least 4 of the 5 intervention steps and those who did not, there was a slightly lower score in the noncompleter group in terms of their agreement with the question on the ease of using the program, and some participants may have discontinued use of the intervention because of technical or usability issues. Although overall the intervention was seen as acceptable to the participants with a good level of engagement observed, it is critical that any usability or acceptability issues are explored in detail and addressed in future redesigns of the intervention. This is important to maximize engagement because the intention to use web-based interventions is mediated by perceived ease of use, usefulness, and social determinants. Increased engagement may be related to behavioral or demographic characteristics, including previous experience of using web-based programs [40,53,54]. In addition, the mode of intervention delivery may have had an influence on engagement, particularly in terms of the number of recorded sessions. The intervention was designed for use on a laptop or desktop computer. These may be accessed less frequently than mobile devices; therefore, delivery of the intervention in a mobile app form might have increased the number of sessions completed by the participants. This was reflected in some of the views on the intervention, with availability of an app format being the second highest suggested improvement that could be made (Table 5).

Use may also have been affected by the behavioral components of the intervention. Although the intervention included common behavior change techniques such as information on health consequences, social support, and use of reminders and prompts, it did not include other methods commonly associated with sustained and repeated use of web-based programs, for example, regular self-monitoring or the use of goal setting. Although the flexibility and open access of the intervention (ie, not locking steps until completion of a previous step) may have increased initial engagement, it might also have been anticipated that this might make it less likely that the participants would return to the intervention as frequently. However, this was not the case, and a reason for this may have been that the users returned to review previous information. This is evidenced by data from the usability survey, which indicated that overall, the intervention was seen as being usable, with the tailored information provided regarded as useful and relevant (Table 4). Another key reported benefit of the intervention was that the participants reported a high level of trust in the information provided and reported that it helped to facilitate them and their partners to initiate conversations about sexual well-being that they might not otherwise have had (Table 4). This was observed despite the participants reporting relatively high levels of comfort in discussing sexual issues with a partner (Table 2). Couple communication about sexual well-being can be regarded as complex and often difficult to initiate [21]. However, such communication is important and is an essential step in managing concerns and supporting sexual well-being recovery.

Limitations

One limitation is that we were unable to explore in detail the reasons for the withdrawal of some participants from the study or examine the factors responsible for increased engagement with the intervention. The sample of participants was also relatively homogeneous, which may limit the generalizability of the study findings. The assessment of usability was based on a modified version of the system usability scale [43], which limits the ability to compare the findings on program usability with those of other studies.

Conclusions

In this paper, efficacy, use, and acceptability data are presented for a tailored web-based intervention designed to maximize sexual well-being in men living with prostate cancer. This study

XSL•FO
RenderX

provides preliminary evidence for the efficacy of the intervention, which was perceived as being usable and acceptable to the participants with evidence of sustained use. Digital interventions may provide access to low-cost, scalable, updatable, and evidence-based information for managing sexual concerns after prostate cancer treatment. By acknowledging the impact of treatment on sexual well-being and providing appropriate support at all stages of care, this intervention might have the potential to improve patient-important outcomes and could easily be made available in routine practice. Further research will be conducted to explore the factors associated with increased engagement, and these findings will be used to refine content before testing as part of larger longitudinal and randomized controlled studies examining longer-term intervention effectiveness on a wider range of patient-important outcomes, including symptom distress, self-efficacy, knowledge, couple communication, sexual satisfaction, and overall quality of life. These studies will also be used to examine the influence of key demographic factors such as age profile, treatment type, relationship status, and sexual orientation on these outcomes.

Acknowledgments

The authors wish to acknowledge their colleague and the principal investigator of this project, the late EMMC. Her research expertise and dedication to those who felt the impact of cancer treatment had a profound impact on this work. The authors would like to thank the Movember Foundation and Prostate Cancer UK for providing funding and support for the TrueNORTH maximizing sexual well-being project that this study is part of. Additional funding was provided by the Health and Social Care Research and Development Division of Northern Ireland. The funders had no role in the study design, data collection and analysis, or preparation of the manuscript. The intervention was developed using the LifeGuide software, which was partly funded by the National Institute for Health Research Southampton Biomedical Research Centre.

Authors' Contributions

EMMC, KP, RM, MS, ST, SJ, and MK developed the original proposal for this study. EMMC, CF, NB, and SROC drafted and revised the additional versions of the intervention content. EMMC, CF, RM, JC, and SROC conducted usability tests and think-aloud interviews during intervention development. SROC, CF, and MS built and tested the final version of the intervention. SROC, CF, and EMMC were responsible for data collection, analysis, and interpretation. SROC drafted the initial manuscript, and all authors revised the manuscript for important intellectual content and approved the final version.

Conflicts of Interest

None declared.

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019 Sep 11;69(5):363-385 [FREE Full text] [doi: 10.3322/caac.21565] [Medline: 31184787]
2. Pishgar P, Ebrahimi E, Saeedi Moghaddam S, Fitzmaurice C, Amini E. Global, regional and national burden of prostate cancer, 1990 to 2015: results from the global burden of disease study 2015. J Urol 2018 May;199(5):1224-1232. [doi: 10.1016/j.juro.2017.10.044] [Medline: 29129779]
3. Donovan JL, Hamdy FC, Lane JA, Mason M, Metcalfe C, Walsh E, et al. Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. N Engl J Med 2016 Oct 13;375(15):1425-1437. [doi: 10.1056/nejmoa1606221]
4. Wong MC, Goggins WB, Wang HH, Fung FD, Leung C, Wong SY, et al. Global incidence and mortality for prostate cancer: analysis of temporal patterns and trends in 36 countries. Eur Urol 2016 Nov;70(5):862-874. [doi: 10.1016/j.eururo.2016.05.043] [Medline: 27289567]
5. Hoffman KE, Penson DF, Zhao Z, Huang L, Conwill R, Laviana AA, et al. Patient-reported outcomes through 5 years for active surveillance, surgery, brachytherapy, or external beam radiation with or without androgen deprivation therapy for localized prostate cancer. J Am Med Assoc 2020 Jan 14;323(2):149-163 [FREE Full text] [doi: 10.1001/jama.2019.20675] [Medline: 31935027]
6. Talvitie A, Ojala H, Tammela T, Koivisto A, Pietilä I. Factors related to self-rated health and life satisfaction one year after radical prostatectomy for localised prostate cancer: a cross-sectional survey. Scand J Caring Sci 2019 Sep 13;33(3):688-697. [doi: 10.1111/scs.12664] [Medline: 30866074]
7. Brundage M, Sydes MR, Parulekar WR, Warde P, Cowan R, Bezjak A, et al. Impact of radiotherapy when added to androgen-deprivation therapy for locally advanced prostate cancer: long-term quality-of-life outcomes from the NCIC CTG PR3/MRC PR07 randomized trial. J Clin Oncol 2015 Jul 1;33(19):2151-2157 [FREE Full text] [doi: 10.1200/JCO.2014.57.8724] [Medline: 26014295]
8. Haahr MK, Azawi NH, Andersen LG, Carlson S, Lund L. A retrospective study of erectile function and use of erectile aids in prostate cancer patients after radical prostatectomy in Denmark. Sex Med 2017 Sep;5(3):e156-e162 [FREE Full text] [doi: 10.1016/j.esxm.2017.06.003] [Medline: 28778680]

XSL•FO
RenderX

9. Downing A, Wright P, Hounsome L, Selby P, Wilding S, Watson E, et al. Quality of life in men living with advanced and localised prostate cancer in the UK: a population-based study. Lancet Oncol 2019 Mar;20(3):436-447. [doi: 10.1016/s1470-2045(18)30780-0]
10. Ross L, Rottmann N, Andersen KK, Hoybye MT, Johansen C, Dalton SO. Distress after a psychosocial cancer rehabilitation course. Main effects and effect modification in a randomised trial at 12 months of follow-up. Acta Oncologica 2015 Mar 9;54(5):735-742. [doi: 10.3109/0284186x.2014.998278]
11. Radhakrishnan A, Henry J, Zhu K, Hawley ST, Hollenbeck BK, Hofer T, et al. Determinants of quality prostate cancer survivorship care across the primary and specialty care interface: Lessons from the Veterans Health Administration. Cancer Med 2019 May;8(5):2686-2702 [FREE Full text] [doi: 10.1002/cam4.2106] [Medline: 30950216]
12. Lee DM, Nazroo J, O'Connor DB, Blake M, Pendleton N. Sexual health and well-being among older men and women in England: findings from the english longitudinal study of ageing. Arch Sex Behav 2016 Jan 27;45(1):133-144. [doi: 10.1007/s10508-014-0465-1] [Medline: 25624001]
13. Speer SA, Tucker SR, McPhillips R, Peters S. The clinical communication and information challenges associated with the psychosexual aspects of prostate cancer treatment. Soc Sci Med 2017 Jul;185:17-26. [doi: 10.1016/j.socscimed.2017.05.011] [Medline: 28549250]
14. Albaugh JA, Sufrin N, Lapin BR, Petkewicz J, Tenfelde S. Life after prostate cancer treatment: a mixed methods study of the experiences of men with sexual dysfunction and their partners. BMC Urol 2017 Jun 15;17(1):45 [FREE Full text] [doi: 10.1186/s12894-017-0231-5] [Medline: 28619019]
15. Skolarus TA, Ragnoni JA, Garlinghouse C, Schafenacker A, Webster D, Hager P, et al. Multilingual self-management resources for prostate cancer survivors and their partners: results of a long-term academic-state health department partnership to promote survivorship care. Urology 2017 Dec;110:92-97. [doi: 10.1016/j.urology.2017.06.017] [Medline: 28652166]
16. Skolarus TA, Wittmann D, Hawley ST. Enhancing prostate cancer survivorship care through self-management. Urol Oncol 2017 Sep;35(9):564-568. [doi: 10.1016/j.urolonc.2017.05.008] [Medline: 28619632]
17. Vartolomei L, Shariat SF, Vartolomei MD. Psychotherapeutic interventions targeting prostate cancer patients: a systematic review of the literature. Eur Urol Oncol 2018 Sep;1(4):283-291. [doi: 10.1016/j.euo.2018.04.011] [Medline: 31100249]
18. Parahoo K, McDonough S, McCaughan E, Noyes J, Semple C, Halstead EJ, et al. Psychosocial interventions for men with prostate cancer. Cochrane Database Syst Rev 2013 Dec 24(12):CD008529. [doi: 10.1002/14651858.CD008529.pub3] [Medline: 24368598]
19. Kirby MG, White ID, Butcher J, Challacombe B, Coe J, Grover L, et al. Development of UK recommendations on treatment for post-surgical erectile dysfunction. Int J Clin Pract 2014 May;68(5):590-608 [FREE Full text] [doi: 10.1111/ijcp.12338] [Medline: 24188207]
20. Prostate Cancer: Diagnosis and Management. National Institute for HealthCare Excellence. URL: https://www.nice.org.uk/guidance/ng131 [accessed 2019-07-01]
21. Mehta A, Pollack CE, Gillespie TW, Duby A, Carter C, Thelen-Perry S, et al. What patients and partners want in interventions that support sexual recovery after prostate cancer treatment: an exploratory convergent mixed methods study. Sex Med 2019 Jun;7(2):184-191 [FREE Full text] [doi: 10.1016/j.esxm.2019.01.002] [Medline: 30833226]
22. Sporn NJ, Smith KB, Pirl WF, Lennes IT, Hyland KA, Park ER. Sexual health communication between cancer survivors and providers: how frequently does it occur and which providers are preferred? Psychooncology 2015 Sep 22;24(9):1167-1173. [doi: 10.1002/pon.3736] [Medline: 25534170]
23. Forbat L, White I, Marshall-Lucette S. Discussing the sexual consequences of treatment in radiotherapy and urology consultations with couples affected by prostate cancer. BJU Int Jan 2012;109(1):98-103. [doi: 10.1111/j.1464-410x.2011.10257.x]
24. Almont T, Farsi F, Krakowski I, El Osta R, Bondil P, Huyghe. Sexual health in cancer: the results of a survey exploring practices, attitudes, knowledge, communication, and professional interactions in oncology healthcare providers. Support Care Cancer 2019 Mar 14;27(3):887-894. [doi: 10.1007/s00520-018-4376-x] [Medline: 30109489]
25. Annerstedt CF, Glasdam S. Nurses' attitudes towards support for and communication about sexual health-a qualitative study from the perspectives of oncological nurses. J Clin Nurs 2019 Oct 20;28(19-20):3556-3566. [doi: 10.1111/jocn.14949] [Medline: 31165516]
26. O'Connor SR, Connaghan J, Maguire R, Kotronoulas G, Flannagan C, Jain S, et al. Healthcare professional perceived barriers and facilitators to discussing sexual wellbeing with patients after diagnosis of chronic illness: A mixed-methods evidence synthesis. Patient Educ Couns 2019 May;102(5):850-863. [doi: 10.1016/j.pec.2018.12.015] [Medline: 30578104]
27. Traa MJ, De Vries J, Roukema JA, Rutten HJ, Den Oudsten BL. The sexual health care needs after colorectal cancer: the view of patients, partners, and health care professionals. Support Care Cancer 2014 Mar 16;22(3):763-772. [doi: 10.1007/s00520-013-2032-z] [Medline: 24240645]
28. Rose D, Ussher J, Perz J. Let's talk about gay sex: gay and bisexual men's sexual communication with healthcare professionals after prostate cancer. Eur J Cancer Care (Engl) 2017 Jan 26;26(1):e12469. [doi: 10.1111/ecc.12469] [Medline: 26918877]
29. Bartels SL, van Knippenberg RJ, Dassen FC, Asaba E, Patomella A, Malinowsky C, et al. A narrative synthesis systematic review of digital self-monitoring interventions for middle-aged and older adults. Internet Interv 2019 Dec;18:100283 [FREE Full text] [doi: 10.1016/j.invent.2019.100283] [Medline: 31890630]

XSL•FO
RenderX

30. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar 9;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
31. Sheeran P, Klein WM, Rothman AJ. Health behavior change: moving from observation to intervention. Annu Rev Psychol 2017 Jan 3;68(1):573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
32. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
33. White ID, Wilson J, Aslet P, Baxter AB, Birtle A, Challacombe B, et al. Development of UK guidance on the management of erectile dysfunction resulting from radical radiotherapy and androgen deprivation therapy for prostate cancer. Int J Clin Pract 2015 Jan 6;69(1):106-123 [FREE Full text] [doi: 10.1111/ijcp.12512] [Medline: 25283500]
34. Turner KM, Rousseau N, Croot L, Duncan E, Yardley L, O'Cathain A, et al. Understanding successful development of complex health and healthcare interventions and its drivers from the perspective of developers and wider stakeholders: an international qualitative interview study. BMJ Open 2019 May 30;9(5):e028756 [FREE Full text] [doi: 10.1136/bmjopen-2018-028756] [Medline: 31152042]
35. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
36. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
37. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
38. Wittmann D, Mehta A, Northouse L, Dunn R, Braun T, Duby A, et al. TrueNTH sexual recovery study protocol: a multi-institutional collaborative approach to developing and testing a web-based intervention for couples coping with the side-effects of prostate cancer treatment in a randomized controlled trial. BMC Cancer 2017 Oct 2;17(1):664 [FREE Full text] [doi: 10.1186/s12885-017-3652-3] [Medline: 28969611]
39. LifeGuide Software. URL: https://www.lifeguideonline.org [accessed 2018-05-01]
40. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. J Med Internet Res 2019 Aug 13;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
41. Chao C. Factors determining the behavioral intention to use mobile learning: an application and extension of the utaut model. Front Psychol 2019;10:1652 [FREE Full text] [doi: 10.3389/fpsyg.2019.01652] [Medline: 31379679]
42. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
43. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interac 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
44. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
45. Connell L, Carey R, de Bruin M, Rothman AJ, Johnston M, Kelly MP, et al. Links between behavior change techniques and mechanisms of action: an expert consensus study. Ann Behav Med 2019 Jul 17;53(8):708-720 [FREE Full text] [doi: 10.1093/abm/kay082] [Medline: 30452535]
46. Knittle K, Heino M, Marques MM, Stenius M, Beattie M, Ehbrecht F, et al. The compendium of self-enactable techniques to change and self-manage motivation and behaviour v.1.0. Nat Hum Behav 2020 Feb 13;4(2):215-223. [doi: 10.1038/s41562-019-0798-9] [Medline: 31932687]
47. Carrier J, Edwards D, Harden J. Men s perceptions of the impact of the physical consequences of a radical prostatectomy on their quality of life. JBI Database of Systematic Reviews and Implementation Reports 2018;16(4):892-972. [doi: 10.11124/jbisrir-2017-003566]
48. Gilbert E, Perz J, Ussher J. Talking about sex with health professionals: the experience of people with cancer and their partners. Eur J Cancer Care (Engl) 2016 Mar 10;25(2):280-293. [doi: 10.1111/ecc.12216] [Medline: 25040442]
49. Bourne S, DeVos R, North M, Chauhan A, Green B, Brown T, et al. Online versus face-to-face pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: randomised controlled trial. BMJ Open 2017 Jul 17;7(7):e014580-e01458- [FREE Full text] [doi: 10.1136/bmjopen-2016-014580] [Medline: 28716786]
50. Kang HS, Kim H, Park SM, Kim J. Online-based interventions for sexual health among individuals with cancer: a systematic review. BMC Health Serv Res 2018 Mar 7;18(1):167 [FREE Full text] [doi: 10.1186/s12913-018-2972-6] [Medline: 29514669]
51. Zhou ES, Bober SL, Nekhlyudov L, Hu JC, Kantoff PW, Recklitis CJ. Physical and emotional health information needs and preferences of long-term prostate cancer survivors. Patient Educ Couns 2016 Dec;99(12):2049-2054 [FREE Full text] [doi: 10.1016/j.pec.2016.07.017] [Medline: 27439668]

XSL•FO
RenderX

52. Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337-e33- [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
53. Breil B, Kremer L, Hennemann S, Apolinário-Hagen J. Acceptance of mhealth apps for self-management among people with hypertension. Stud Health Technol Inform 2019 Sep 3;267:282-288. [doi: 10.3233/SHTI190839] [Medline: 31483283]
54. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb 20;15(2):e37-e3- [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by I Wilson, G Dowsett, PhD; comments to author 01.09.20; revised version received 16.09.20; accepted 11.03.21; published 26.07.21.

Please cite as:
O'Connor SR, Flannagan C, Parahoo K, Steele M, Thompson S, Jain S, Kirby M, Brady N, Maguire R, Connaghan J, McCaughan EM
Efficacy, Use, and Acceptability of a Web-Based Self-management Intervention Designed to Maximize Sexual Well-being in Men Living With Prostate Cancer: Single-Arm Experimental Study
J Med Internet Res 2021;23(7):e21502
URL: https://www.jmir.org/2021/7/e21502
doi:10.2196/21502
PMID:34309580

©Sean R O'Connor, Carrie Flannagan, Kader Parahoo, Mary Steele, Samantha Thompson, Suneil Jain, Michael Kirby, Nuala Brady, Roma Maguire, John Connaghan, Eilis M McCaughan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Older Adults' Experiences of Behavior Change Support in a Digital Fall Prevention Exercise Program: Qualitative Study Framed by the Self-determination Theory

Beatrice Pettersson[1], MSci; Rebecka Janols[2,3], PhD; Maria Wiklund[1], PhD; Lillemor Lundin-Olsson[1], PhD; Marlene Sandlund[1], PhD

[1]Physiotherapy, Department of Community Medicine and Rehabilitation, Umeå, Sweden

[2]Occupational Therapy, Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden

[3]Department of Computing Science, Umeå University, Umeå, Sweden

Corresponding Author:
Beatrice Pettersson, MSci
Physiotherapy, Department of Community Medicine and Rehabilitation
Umeå University
Umeå, 90187
Sweden
Phone: 46 907865224
Email: beatrice.pettersson@umu.se

Abstract

Background: Exercise is an effective intervention to prevent falls in older adults; however, long-term adherence is often poor. To increase adherence, additional support for behavior change has been advocated. However, consistency in the reporting of interventions using behavior change techniques is lacking. Recently, a classification system has been developed to increase consistency in studies using behavior change techniques within the self-determination theory.

Objective: This study aimed to explore expressions of self-determination among community-dwelling older adults using a self-managed digital fall prevention exercise program comprising behavior change support (the Safe Step program), which was developed in co-creation with intended users.

Methods: The qualitative study design was based on open-ended responses to questionnaires, and individual and focus group interviews. A deductive qualitative content analysis was applied using the classification system of motivation and behavior change techniques as an analytical matrix, followed by an inductive analysis. Twenty-five participants took part in a feasibility study and exercised in their homes with the Safe Step program for 4 months. The exercise program was available on computers, smartphones, and tablets, and was fully self-managed.

Results: In the deductive analysis, expressions of support were demonstrated for all three basic human psychological needs, namely, autonomy, competence, and relatedness. These expressions were related to 11 of the 21 motivation and behavior change techniques in the classification system. The inductive analysis indicated that autonomy (to be in control) was valued and enabled individual adaptations according to different rationales for realizing exercise goals. However, the experience of autonomy was also two-sided and depended on the participants' competence in exercise and the use of technology. The clarity of the program and exercise videos was seen as key for support in performance and competent choices. Although augmented techniques for social support were requested, support through relatedness was found within the program.

Conclusions: In this study, the Safe Step program supported the establishment of new exercise routines, as well as the three basic human psychological needs, with autonomy and competence being expressed as central in this context. Based on the participants' experiences, a proposed addition to the classification system used as an analytical matrix has been presented.

Trial Registration: ClinicalTrials.gov NCT02916849; https://clinicaltrials.gov/ct2/show/NCT02916849

(J Med Internet Res 2021;23(7):e26235) doi:10.2196/26235

KEYWORDS

accidental falls; aged; exercise; qualitative research; eHealth; self-management; fall prevention; behavior change; self-determination theory; classification of motivation and behavior change techniques

XSL•FO
RenderX

Introduction

Background

Fall prevention interventions for older adults living in the community have the potential to reduce the physical and psychological distress resulting from a fall [1]. A recent meta-analysis showed that physical exercise, in particular, is an effective intervention for the reduction of fall rates and number of falls [2]. However, fall prevention exercise interventions have failed to produce lasting long-term results owing to a lack of successful behavior change and sustained exercise routines among participants [3,4]. To better understand older adults' exercise adherence and health behavior, the research community has identified a need to increase the use of behavioral scientific knowledge in the development, use, evaluation, and reporting of interventions [5-7]. However, interventions aimed at supporting increased physical activity and improving adherence to interventions among older adults are rarely underpinned by behavior change theories [5].

There is growing interest in using digital technology to support behavior change, increase access to fall prevention interventions, and promote the self-management of physical activity [8-10]. Older adults have expressed an appreciation for the increased autonomy and flexibility provided by digital programs, which allow them to exercise at a time that suits them in a location of their choosing [11-13]. Moreover, the incorporation of individual and social behavioral change components in digital exercise programs has proven more successful in improving adherence and physical performance among older adults in comparison with those receiving a home exercise booklet [14]. There is still a great need to further investigate the feasibility of self-managed digital exercise interventions for older adults. Notably, a systematic review indicated that techniques that increase self-efficacy and physical activity behavior in a younger population are not as effective in adults aged 60 years or older [15]. Consequently, involving end-users in the design and development of new programs could be a way to increase the adoption of health-related digital interventions among older adults. It is of major importance that the interventions are reported in a structured way and evaluated in terms of which behavior change techniques older adults find supportive and in which context.

The aim of this study was to explore expressions of self-determination among community-dwelling older adults using a self-managed digital exercise program for fall prevention with co-created behavior change support, and to explore a classification system based on self-determination theory (SDT) as an analytical matrix in this self-managed digital context.

SDT and Classification of Motivation and Behavior Change Techniques

SDT points to the individual's perceived self-regulation and internalization process as an explanatory factor in behavioral outcomes [16]. Perceived self-regulation is based on the "satisfaction" or "frustration" of the following three basic human psychological needs: *autonomy*, *competence*, and *relatedness* [17]. Autonomy represents the feeling of self-endorsement or volition in one's behavior. Competence represents a feeling of capability and reflects an ambition to perform and improve skills. Lastly, relatedness represents being a significant part of a social context, which evokes feelings of being connected and valued. Autonomy is central to SDT, as it is through this construct that the other basic psychological needs are actualized. Therefore, the needs determine the quality of motivation, but they are dependent on the level of autonomy [18]. According to SDT, all three basic human psychological needs have to be satisfied in order to achieve self-motivation and experience well-being [16].

It has been concluded that techniques used within SDT-based interventions cannot fully be captured by existing taxonomies [19,20]. Therefore, the classification of motivation and behavior change techniques (MBCTs) was recently developed in an expert consensus by Teixeira et al [20]. The purpose was to increase consistency in the identification and reporting of techniques between studies using SDT in health contexts. The classification comprises 21 MBCTs organized according to the techniques' support for each psychological need, accompanied by labels, definitions, and function descriptions. An MBCT is defined as "a distinct, observable, and replicable component of an intervention, designed to influence a person's behavior directly or indirectly by impacting the person's perceptions of autonomy, relatedness, and/or competence need satisfaction in relation to a particular behavior or group of related behaviors" [20].

Methods

Study Setting and Design

This study is part of a larger project with the overarching aim to develop and evaluate self-managed digital fall prevention for older adults. An overview of the whole project and its parts is presented in Multimedia Appendix 1.

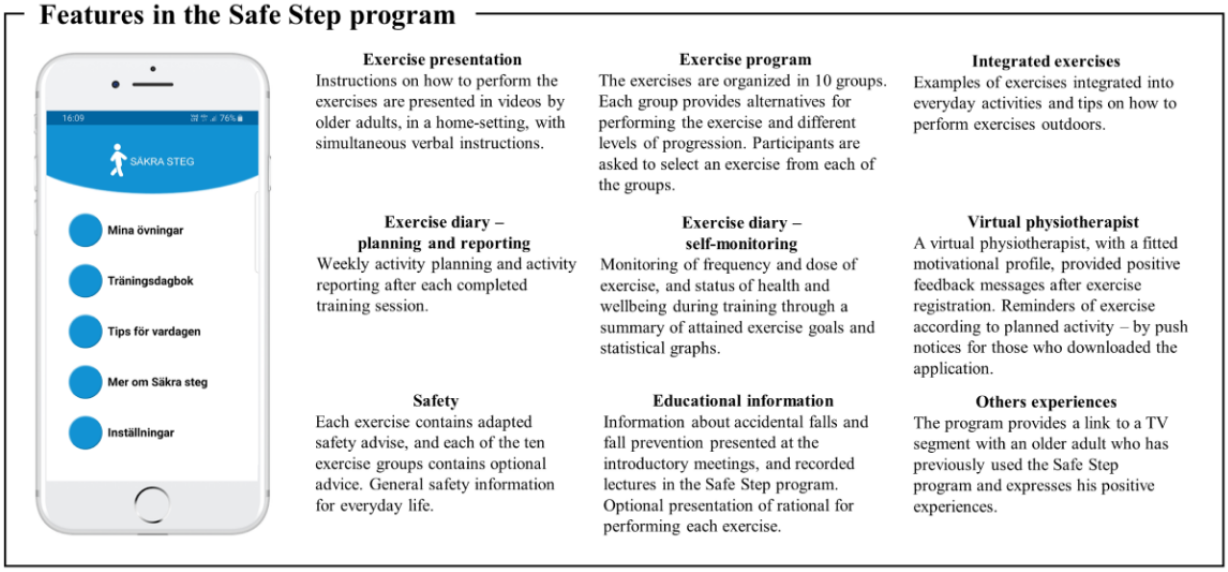
Together with older adults, we have co-created behavior change support for a digital self-managed exercise program. The behavior change components were integrated in an application and tested in a feasibility study for 4 months. This qualitative study involved participants using this digital program, and it presents the findings of individual and focus group interviews, as well as open-ended responses to questionnaires, all of which were part of the feasibility study conducted in Umeå, Sweden (ClinicalTrials.gov NCT02916849) between September 2016 and February 2017. The feasibility study compared two home-based self-management exercise programs (a digital program [Safe Step v.1] and a paper booklet) [21]. Participants who enrolled in the study could choose the digital program or the paper booklet. In all, 29 participants chose the digital program. Deductive qualitative content analysis was used to explore the older adults' experiences of the behavior change support. The classification of MBCTs served as an analytical matrix [20]. Further explorations were made through inductive qualitative content analysis [22]. The Consolidated Criteria for Reporting Qualitative Research (COREQ) [23] were used to guide the reporting.

XSL•FO
RenderX

Development of the Program and Behavior Change Support

In two steps, the Safe Step program and the behavior change support evaluated in this study were developed by researchers from Umeå University and Luleå Technical University together with older adults, with the majority of external funding coming from government funding bodies. The development of the program was facilitated by the method Participatory and Appreciative Action and Reflection [24]. An overview of features developed in the program is presented in Figure 1. In the first step, a self-management digital exercise program, consisting of, for example, exercise presentation, an exercise program, and integrated exercises, was developed by a multidisciplinary team of researchers in collaboration with older adults [25]. In the second step, through further co-creation, specific behavior change techniques, that is, an exercise diary including planning, reporting, and self-monitoring, as well as a virtual physical therapist, were jointly developed by older adults and researchers with competence in physical therapy and usability.

Figure 1. Overview of features in the Safe Step program.



The Safe Step Digital Exercise Program

The Safe Step program was available on computers, smartphones, and tablets, and was free of charge. An introduction meeting provided a short introduction to fall prevention and how to navigate and use the digital program. The program was fully self-managed, that is, participants were asked to plan, execute, and evaluate the training independently. As part of the self-management of exercise, the participants were expected to compose their own program of 10 exercises from predetermined groups of exercises with a different focus as follows: lower-limb muscle, balance, and gait/step. The exercises were presented through videos, where older adults performed the exercises in a home setting. All participants were asked to exercise at least 30 minutes, 3 days a week, for 4 months. As part of the feasibility study, participants were offered to participate in a peer-mentoring meeting once a month, which one-third of the participants chose to do.

Data Collection

Data for descriptive purposes were collected at baseline through a questionnaire containing demographic factors, previous use of technology, and medical history. Data analyzed in this study were collected at two timepoints. Questionnaires were answered 2 months after study start, and focus group interviews and individual interviews were conducted at the end of the 4-month intervention study. An overview of the focus areas for the questions asked on the different data collection occasions is presented in Multimedia Appendix 2. Among the three data collections, 10 participants contributed in one data collection, eight participants contributed in two data collections, and seven participants contributed in all three. In total, 25 participants contributed in the data collection. Their mean exercise time with the program per week was 78 minutes (minimum 11 minutes, maximum 158 minutes).

Questionnaires

At 2 months, three participants had withdrawn from the study. The remaining 26 participants received questionnaires, which were sent by mail and contained a postage-paid envelope for the return of the questionnaires. Questions regarding their experiences using the Safe Step program were asked, particularly relating to the individual exercise program, the ease of using of the application, and experiences of the different features, such as planning their weekly exercise and the virtual physical therapist. Each of the questions was accompanied by a text field in which the participants were encouraged to write explanatory comments. Of the 26 returned questionnaires, 18 had comments, which were included as units of analysis in this study. The comments varied from a few words to extensive comments.

Focus Group Interviews

At 4 months, two more participants had withdrawn from the study. A purposely selected sample of 14 participants was sent an invitation to participate in focus group interviews. We aimed for a representation of both men and women who had used the program on computers, smartphones, and tablets, as well as variation of exercise frequency. The participants could choose

XSL•FO
RenderX

between two different dates to attend the focus group interviews, which were held in a conference room at the university. One woman who signed up for focus group 1 and one man who signed up for focus group 2 withdrew before the interviews began, which resulted in a more unequal distribution of gender than intended. Two moderators conducted the interviews (BP, a physical therapist; RJ, a usability expert), both with previous experience conducting focus group interviews. The interviews were 92 and 87 minutes in length and were facilitated by a focus group guide with open-ended questions (Multimedia Appendix 2). The questions focused on the participants' experiences of the behavior change techniques developed in the second part of the program development, that is, the exercise diary, the virtual physical therapist, feedback messages and reminders, the approach to make a weekly plan for the exercises, and self-monitoring features. The initiation question concerned the experience of exercising with a program in a digital format.

Individual Interviews

A purposive selection of 17 participants, out of the 24 who had completed the digital exercise program, were asked to participate in individual semistructured interviews by a member of the research team at the postassessment of the feasibility study. In order to obtain a varied selection of participants, selection was made based on gender, place of recruitment, and exercise adherence. One of the interviews was conducted with a married couple who both used the digital program. According to the participants' preference, the interviews were held at the university, at home, or at a library. An interview guide with open-ended questions facilitated the interviews (Multimedia Appendix 2). The questions were related to the participants' experience of self-managing their exercise and included those regarding the composition of the individual program, perceived effects of the exercise, strategies for and feelings of safety while performing the exercise, strategies and support for maintenance of the exercise, structure and content of the program, use of the program (when, where, and how), and recording in the exercise diary. Follow-up questions were asked, and the participants were encouraged to give concrete examples of their experiences. The interviews were performed by a physical therapist with extensive experience conducting interviews or by one of two physical therapy students with experience conducting interviews. These interviews have previously been analyzed in a study, but with another aim and with both intervention groups included [11]. Background characteristics of the participants are presented in Table 1.

Table 1. Participant characteristics.

Characteristic	Questionnaires (n=18)	Focus group 1 (n=6)	Focus group 2 (n=6)	Individual interviews (n=17)
Age (years), mean (min-max)	76.5 (71-91)	76 (72-79)	74 (71-79)	76 (71-91)
Women, n (%)	9 (50%)	1 (17%)	4 (67%)	10 (59%)
Access to a smartphone/tablet, n (%)	14 (78%)	3 (50%)	4 (67%)	14 (82%)
Access to a computer, n (%)	15 (83%)	6 (100%)	5 (83%)	14 (82%)
Household situation, n (%)				
Living together	13 (72%)	3 (50%)	4 (67%)	12 (71%)
Living alone	5 (28%)	3 (50%)	2 (33%)	5 (29%)

Data Analysis

All interviews were audio-recorded and transcribed verbatim. The data analysis was performed using qualitative content analysis in two stages (deductively and then inductively) [26]. Qualitative content analysis is suitable for exploring variations within data as it highlights differences and similarities [22,27]. The deductive (theory-driven) approach enables exploration of previous knowledge in a different context [26], whereas the inductive (data-driven) approach generates new knowledge by searching for patterns in the data and moving toward a more abstract level of understanding [28].

The classification of MBCTs was used as a categorization matrix in the deductive analysis [20] with the 21 MBCTs as predetermined categories. To capture experiences of self-determination not addressed by MBCTs, the three basic psychological needs tied to SDT [16] were used as main categories in the categorization matrix. Initially, the first author (BP) performed a read through to get a sense for the entirety of the material. Units relevant to the aim of this study were selected and condensed, and representative codes were created. The codes were simultaneously sorted in the categorization matrix. Following the deductive stage, an inductive stage was initiated to compare the codes for similarities and differences and form subcategories within each category of the classification of MBCTs. Data analysis was performed using the software MAXQDA 2020 (VERBI Software). In the context of this study, 10 MBCTs were found to be nonrepresentative. Moreover, codes perceived as corresponding to psychological needs, but not exemplified in the classification system, served as a basis for an additional competence-based category. The analysis was continuously discussed and agreed upon by the authors BP, MW, LLO, and MS. To ensure trustworthiness, these authors also deductively analyzed one individual interview and discussed their interpretations [22]. The emergent results were discussed with the second author (RJ) who was also part of data collection.

Results

Overview

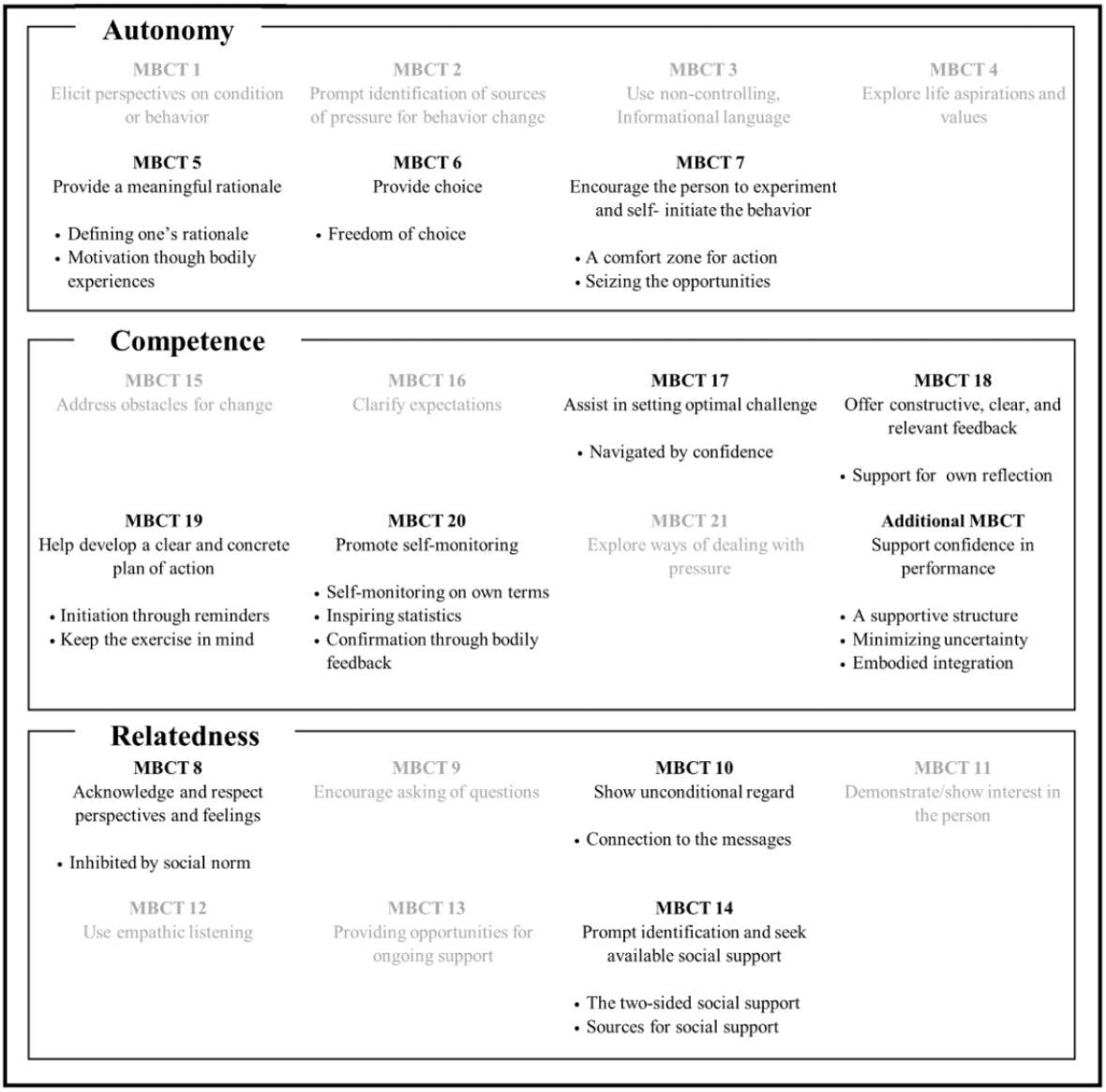
Participants' expressions of self-determination showed that the Safe Step digital exercise program was perceived as a supportive structure that guided their training, providing a feeling of freedom of choice and ownership over their courses of actions.

XSL•FO
RenderX

With increased confidence in their actions, the participants were able to find their own meaningful and motivating ways to work with the program. Still, the digital exercise program could entail both freedom and restrictions due to differences in confidence in managing the program and varying needs for social connectedness.

As a result of the deductive analysis, we found expressions for all three psychological needs in the SDT [16] (the main categories). Expressions of autonomy and competence were found to be more central than relatedness. Furthermore, the expressions corresponded to 11 of the 21 MBCTs (the categories), and an additional competence-based category was created. In turn, the inductive analysis generated 19 subcategories within these categories. The results are presented in an overview in Figure 2.

Figure 2. Overview of the motivation and behavior change techniques (MBCTs) according to the classification of MBCTs, where MBCTs found nonrepresentative of our study are presented in grey. The basic human psychological needs are presented in the order in which the participants expressed the most support. The bullets represent subcategories.



Autonomy Supportive Techniques

The autonomy supportive categories represented the core of the participants' expressions of support for changes in exercise behavior. Choices regarding exercises and exercise routines were a prerequisite for feeling ownership over their actions and meaningfulness in their performance. Thus, "satisfaction" of autonomy was primarily conveyed.

Provide a Meaningful Rationale (MBCT 5)

The identification of a meaningful rationale for performing the exercises was captured in the subcategories "Defining one's rationale" and "Motivation through bodily experiences."

The digital program either supported existing rationales for doing the exercises or helped participants develop new rationales. The motivation for doing the exercises was often expressed in terms of long-term goals, such as strengthening leg muscles or being able to dress while standing. Participants

XSL•FO
RenderX

also reported a fear of deteriorated psychological and physical health, and the exercises were perceived as a tool to counteract this type of decline and maintain improvements.

> *This has been good for me, so I cannot suddenly stop. I might get my problems back then.* [Woman 1; individual interview]

Participants also conveyed an awareness of the economic burden of falls on society. This knowledge had been gained in the introduction of the study; therefore, this had also become a meaningful rationale for doing the exercises. Another rationale and further motivation involved experiences of bodily improvements, which confirmed the effectiveness of the exercise performed. These meaningful improvements of strength and balance were noted not only in everyday activities, such as walking up the stairs, but also when performing the Safe Step exercises. Performing the exercises acted as an eye opener for their issues with balance, which served as a motivator to continue doing the exercise. One participant made the following comment:

> *When it gets easier to do, it becomes fun, and then enjoyable.* [Woman 2; individual interview]

Provide Choice (MBCT 6)

Different aspects of ownership over the course of action through choices were highlighted in the subcategory "Freedom of choice."

Freedom of choice was obtained by being able to choose both a time and place for exercise, what device the program would be used on, and which program features to use. The program was utilized according to participants' own wishes, which created a feeling of ownership over their choices of actions. The routines participants developed were not seen as sustainable if they were not able to find the right arrangement for themselves, put as follows:

> *There has to be choices for all the different ways of thinking.* [woman; focus group interview 2]

The wide range of exercises within the individual training program was seen as a decisive factor in finding the right arrangement for themselves. Conversely, the need to create their own exercise program could also be perceived as a laborious process and a level of independence they did not find desirable.

> *The big advantage is that you get the videos in the Safe Step program, but it's difficult when you have to sit and watch them all to be able to choose.* [Man 1; individual interview]

Encourage the Person to Experiment and Self-initiate the Behavior (MBCT 7)

The subcategories "A comfort zone for action" and "Seizing the opportunities" capture expressions of how participants implemented the strategies and proposals given in the program and found ways to do the exercises that worked with their routines and motivation.

Comfort zones for performing the exercise were developed from suggestions in the program and described as enjoyment experimenting or stick to a structured training routine. Some participants found experimentation to be an enjoyable aspect, for example, by performing the exercises in outdoor environments and integrating them into their everyday routines. Others expressed their preference for the comfort of a familiar routine, and the commitment to complete the requested exercise amount could be perceived as a relatively high commitment. Another aspect of experimentation with the program, beyond the suggestions given to the participants in the program, was spontaneous attention given to capturing opportunities for exercise, both in terms of time and place for practice. One conversation was as follows (focus group interview 2):

> *Man 1: I really cracked the code of doing extra exercises.*
>
> *Woman 1: With the toes?*
>
> *Man 1: No, all kinds of situations. When I walk around the apartment, it suddenly hits me, and I start walking on my heels.*

The ability to seize opportunities was described as something that was learned and evolved over time.

Taken as a whole, the autonomy-based categories highlight the importance and desire for opportunities for choice and personalization when using the digital exercise program.

Competence Supportive Techniques

The categories created from the data analysis demonstrated that competence was seen as something that was developed through supported confidence and a strengthened inner locus of causality. Steps were taken to tailor the program to individual needs, rationales, and life circumstances, thus contributing to the "satisfaction" of the need for competence.

Assist in Setting an Optimal Challenge (MBCT 17)

The participants experienced support to set a suitable challenge, and the confidence and prerequisites for doing so were represented in the subcategory "Navigated by confidence."

The digital exercise program was expressed as support for implementation of new exercise routines and assisted in setting weekly exercise goals from suggestions provided by the program. Variations in confidence in the participants' own ability to achieve an optimal challenge were noted. For those with greater confidence, the suggestions provided in the exercise videos were enough to give a sense of capability to find a suitable exercise intensity and enabled them to adapt quickly, for example, to pain, by switching to easier or more difficult exercises.

> *I have always adapted to what suits me best because of my pain.* [woman; focus group interview 2]

Some expressed confidence in progressing their exercise but felt restricted due to limited progression options for their capability. Self-managing exercise progression could also be experienced as a challenge due to fear of having difficulties with the technology, voiced as follows:

> *I haven't tried to change exercises because I worried there would be trouble with the equipment. Maybe it would not work.* [Man 2; individual interview]

XSL•FO
RenderX

Offer Constructive, Clear, and Relevant Feedback (MBCT 18)

The importance of nuanced and personalized feedback to support reflection on training performance was emphasized in the subcategory "Support for own reflection."

The in-application statistics were communicated as clear and supportive for own reflections. Several benefits of being able to track changes over time were presented, and they were voiced as follows:

> *You get a grip of the situation.* [Man 3; questionnaire]
>
> *Statistics are always revealing.* [Man 1; questionnaire]

To be able to convey causes to adaptions through notes was expressed as a basis for being able to do a better self-evaluation of the training period. Further, the content of the feedback messages sent by the virtual physical therapist could either be perceived as supportive and a trigger for reflection, or lacking meaningful tailored feedback.

> *I like to go back and see which exercise I did not do when it [the virtual physical therapist] says I did not do them all. But, they appear when I have already done my training and I am not really interested to read much then.* [Man; focus group interview 1]

There was thus both "satisfaction" and "frustration" regarding support for own reflections.

Help Develop a Clear and Concrete Plan of Action (MBCT 19)

The subcategories "Initiation through reminders" and "Keep the exercise in mind" demonstrated feelings of support to develop new training routines.

Receiving reminders to support the initiation of the training generated diverse opinions. It was expressed that to do the exercise, the thought of exercise needed to be kindled, and for some, the reminders served that purpose. However, the lack of choices regarding when to receive them and how often was perceived as frustrating. By this, the need for tailored options was stressed. One conversation was as follows (focus group interview 2):

> *Woman 1: I would still like to be able to receive reminders.*
>
> *Woman 2: Yes, preferably a few, but not too many. Just enough would be the best.*

The weekly planning was highlighted as providing a valuable structure and a learning instrument to develop individual exercise routines. The importance of flexible planning was strongly emphasized as follows:

> *Weekly planning should guide training and not be considered the law* [Man; focus group interview 1]
>
> *The plan is a goal, but reality controls when exercise is suitable.* [Man 1; questionnaire]

Promote Self-monitoring (MBCT 20)

The subcategories "Self-monitoring on own terms," "Inspiring statistics," and "Confirmation through bodily feedback" represented that self-monitoring was governed by personal preferences, individual needs, and bodily knowledge.

Self-monitoring of training was expressed to be on own terms. Overall, registering exercise was expressed to work well and became part of the workout routine. However, after a while, some of the participants chose to stop registering or created their own systems for taking notes of performance. Nonetheless, it was emphasized that registration supported development of routines.

> *Even though I appreciate to register my exercise, I exercise for my own sake, not anyone else's, or the registrations.* [Woman 6; individual interview]

The in-application statistics were also expressed to encourage self-monitoring of performance. Informative overviews helped guide exercise management and acted as motivators to continue to follow progress. As encouraged in the exercise videos, bodily feedback informed decisions about the individual exercise program. Experiences of improvements or declines in physical status, such as fatigue and improved or declined strength or steadiness, augmented self-monitoring and contributed with guidance in the adaptations of exercise intensity, voiced as follows:

> *I felt really insecure when I was walking in the city center. Now when I am out walking I can actually look at my surroundings, which are actually really pretty [laughs].* [Woman 6; individual interview]

Support Confidence in Performance (Additional MBCT)

The subcategories "A supportive structure," "Minimizing uncertainty," and "Embodied integration" referred to feelings of ease of use of the program and developed confidence in performing the exercise.

To be able to initiate a new exercise routine, the importance of a supportive structure was stressed for easily getting started, as was ease of use of the digital program, stated as follows:

> *You should not be able to misinterpret anything.* [Woman 5; individual interview]

These important qualities translated to the experience of the Safe Step program. Nonetheless, needs were expressed of a more extensive introduction on how to manage the program. Key elements of minimizing uncertainty regarding how to perform the movements were being able to see the exercises in video format and simultaneously being able to hear the instructions, which were also seen as access to ongoing support.

> *I think video tutorials is much better than a picture with text. Moving instructions along with telling me how it is performed, you can't misunderstand that.* [Man 4; individual interview]

The clarity of the instructions was mentioned as a basis for gradual integration and transition from needing the support of videos to be able to perform the exercises, to continuing without support. The exercise repetitions created an embodied knowledge and integration, which could be experienced as follows:

XSL•FO
RenderX

> *If you have done the exercises for a long time, they are in your head. Now when I do them without the tablet I almost hear them speak.* [Woman; focus group interview 2]

Through a comprehensive understanding of the program, they were able to easily integrate relevant parts of the program at any given situation and thereby find new situations for practice.

Taken together, competence to self-manage exercise was supported by the clear structure of the digital program and guided by confidence to manage the program.

Relatedness Supportive Techniques

The categories show the diversity of the participants' feelings and needs for social support while exercising with the digital program. In this sense, the basic psychological need of relatedness was both "satisfied" and "frustrated." Aspects both in the close use of the program and by age-related perceptions of exercise influenced experiences of relatedness.

Acknowledge and Respect Perspectives and Feelings (MBCT 8)

Emotions that arose while interacting with the environment during performance of the training is described in the subcategory "Inhibited by social norm."

As expressed by the participants, choices made in the program were influenced by a need to comply with the expected social norm of how to perform physical activity. The need to conform to such a norm became a contributing factor to which exercises were selected, and influenced where the training felt comfortable to perform.

> *Jumping steps felt a bit silly to do around the apartment. It is quite close between the apartment buildings and you can see the room quite clearly from the other building (laughter).* [Man 5; individual interview]

Performing the exercises outdoors could be experienced as uncomfortable at first, but individuals could gradually grow accustomed to it.

Show Unconditional Regard (MBCT 10)

The subcategory "Connection to the messages" portrayed thoughts of the support given by the feedback messages.

The information received from the virtual physical therapist was either perceived as an encouraging influence or an unimportant element. The encouraging influence was substantiated by an appreciated pleasant and praiseworthy tone of the messages. It was valued that someone took an interest in their performance, and the messages were interpreted as nonautomatically generated.

> *It felt like someone really cared about that I participated.* [Man; focus group interview 2]

By others, the messages could be experienced as an unimportant element due to a feeling of impersonal content, expressed as follows:

> *Receiving the messages just became part of the routine.* [Man; focus group interview 2]

No interest was sparked, and as a consequence, the messages were read at the beginning, but gradually the interest was lost.

Prompt Identification and Seek Available Social Support (MBCT 14)

The subcategories "The two-sided social support" and "Sources for social support" summarized different needs and solutions for social cohesion and social support to facilitate a changed exercise behavior.

The interviews revealed the two-sided nature of the need for social support. On the one hand, a greater freedom of choice was expressed when self-managing exercise at home, which could also be time-saving, and voiced as follows:

> *I have had to be on time my whole life. I don't want to do that anymore.* [Woman 2; individual interview]

On the other hand, some missed the social support and felt that a group context would be more enjoyable and help increase the frequency and intensity of the exercise. Exercising at home was communicated as requiring a particularly large commitment.

> *Too easily it happens that the one sitting on your shoulder says: not today, and not tomorrow either, so the training does not happen.* [Woman 1; individual interview]

Finding own ways to increase social support during training and to identify sources of support was described. The program could act as a tool for social interaction as participants sometimes shared their knowledge or invited others, in whom they had identified a need for exercise, to exercise with them.

> *When I visit my friends, I have almost overwhelmed them when I surprise them with: "Now, let's do some balance exercises", and they say: "Have you gone crazy?" But, then we perform them, and they are no better than me. It has become a fun event.* [Woman 2; individual interview]

Several sources of social connection were addressed within the application. The older adults demonstrating the exercises in the exercise videos had for some become virtual "friends" to keep them company during exercise, and the voice giving instructions in the videos was also perceived as a pleasant company. The virtual physical therapist was spoken of in terms of him or her, but the content of the messages was perceived as the most important element and not the avatar.

In summary, the above categories make it clear that even with a fully self-managed digital program, ways to increase relatedness were found, both interpersonally and through the digital program. A desire for additional alternatives for support and social cohesion was communicated, as was an awareness and adaptation to social norms.

Discussion

Principal Findings

This study explored expressions of self-determination among older adults using a self-managed digital exercise program with behavior change support (the Safe Step program). The results demonstrate that the Safe Step program was experienced as

XSL•FO
RenderX

providing a supportive structure and an opportunity to tailor the individual exercise program, which was appreciated, although the freedom of choice was sometimes experienced as challenging. Support for all three basic psychological needs when using the Safe Step program was expressed, but autonomy and competence were more central to the participants' experiences than relatedness, and as such, they were identified as primarily "satisfied." The basic psychological need of relatedness was appreciated when present, but was requested by some participants, and therefore, it was identified as somewhat "frustrated."

Fundamental to the participants' expressions of autonomy was the experienced freedom of choice when using the digital program, as well as the possibility to adapt the exercise routines to their personal preferences and circumstances. Offering the opportunity to make informed and reflective choices is considered supportive for autonomy [16] and strengthening for behavioral engagement [29]. In accordance with our findings, a qualitative study involving older adults and health care professionals suggested that digital technology should enable participant independence and self-control in changing physical activity behavior and additionally should be perceived as nondemanding [30]. Further, older adults' rationales, preferences, and possibilities to influence their fall prevention exercises have been found to be important to support a sustained intervention engagement [31].

When using digital technology, intervention engagement has been suggested to be affected by need satisfaction in different spheres as follows: adoption of new technology; interaction with the interface; engagement with a technology-enabled task; technology-supported behavior; link between technology and overall well-being; and societal impact [7]. As such, autonomy is experienced not only in the spheres of the interface or task through ease of use and choices, but also by extension through increased volition, removal of obstacles, or augmentation of capabilities in matters of daily life [7]. Increased autonomy in daily life was portrayed as an intervention effect in our study through increased confidence to experiment with the training and better capabilities to perform everyday tasks challenging balance and strength.

Support for competence to self-manage exercise was expressed as developed from a clear structure and clarified expectations, although a sustained lack of confidence to manage the digital program was perceived by some. Together with older adults, we developed techniques supportive for self-management of exercise. As presented in the competence-based categories, they were overall perceived as supportive for developing personalized exercise routines and confidence in performing them. Taken together, the Safe Step program supported a sense of capability, increasing the sense of self-efficacy for exercises, which was also found in a previous qualitative study [11]. In contrast, in other studies, behavior change techniques associated with self-regulation have been found less supportive for exercise self-efficacy among older adults [15]. However, the more positive attitude presented in our study could be related to older adults' awareness of the self-regulatory nature of the intervention at enrollment.

In this study, the visual guidance in the exercise videos was highlighted as especially supportive for feeling confident in exercise performance. In concurrence with our results, demonstration of exercises has previously been suggested as helpful in supporting older adults' changed physical activity behavior [15,30]. Moreover, in the Behavior Change Technique Taxonomy v1 (BCTTv1) [32] "Demonstration of the behavior" and "Instructions on how to perform the behavior" are specific behavior change techniques. In a systematic review and meta-analysis of SDT strategies used in health interventions, these techniques have been suggested to relate to the strategy of structure. Structure was defined as follows: "practitioners set parameters within which choice and agency can take place and provide support to initiate action" [19]. Structure has been suggested to be both autonomy [19] and competence supportive [33]. This overlap is, however, not uncommon as strategies often can be supportive for more than one need [16]. In the classification of MBCT, we found no support for classifying expressions related to how the program provided structure and supported competence in the actual performance of the exercise. Hence, an additional competence-based category was created as the experiences expressed by the participants were related to increased confidence in performing the activity and support for improving skills.

A wish for increased relatedness in connection to training or a belief of additional benefits of group-based exercise was expressed in this study. This preference for group training has also been expressed in previous studies as a way to make the exercise delivered by digital technology more enjoyable [30]. A way to support relatedness in digitally supported home-based training could be to include tips on how the exercises can be performed with others. Interestingly, some participants had already started to use the application in their interaction with others, indicating both a pride in their efforts and a confidence in their capability to guide others. Apart from using the program as a facilitator of relatedness, connectedness to the digital program itself was conveyed in the interviews. Similar results presented in previous research regarding relatedness and telecare suggest that the concept of relatedness should be broadened beyond the interpersonal context to incorporate organizational and technical concerns in order to provide a more nuanced view [34]. It is worth noting that the importance of relatedness has previously shown inconsistencies in training contexts, which may be partly due to variations in the design and measurements of the studies [35,36]. Still, as also mentioned in this study, individuals might prefer to undertake fall prevention exercise in solitude [37], which may reduce the importance of supporting relatedness in that activity in contrast to autonomy and competence [18].

We found that participants sometimes avoided certain public settings or exercises due to an awareness of sociocultural ageist norms for exercise and the aging body [38]. Previous qualitative results illustrated that negative age perceptions for physical activity can reside within the individual and must be overcome to facilitate sustained exercise behavior. Furthermore, older adults felt that they were treated differently because of their age when exercising in public spaces or social situations [39]. As part of counteracting such ageist norms, the exercise videos in

XSL•FO
RenderX

the Safe Step program were filmed with older adults. However, our results indicate that further efforts could be made. One development could be to incorporate more films with other participants who voice emotions and problematize or even challenge restricting age norms.

Satisfaction of the basic psychological needs has previously been found to support exercise behavior among older adults. However, in likeness with our results, the representativeness of the different needs varied [35,40]. Results from a meta-analysis showed that no individual SDT-based technique to promote motivation for health behavior change was predominant, and a need supportive environment should contain multiple techniques to promote health behavior change [19]. These results are comparable to the results of our study, as the participants' experiences were more shaped by the sum of behavior change support than by the parts.

Theoretical and Methodological Discussion

In our previous research where older adults' exercise preferences and motivators in the context of fall prevention were explored, it emerged that SDT captured the essence of the findings well [41]. Accordingly, during the developmental phase of Safe Step, the SDT did influence the content and layout, but the theory was not explicitly used to frame the studies or the interview guides. Because of the theory's influence, the classification of MBCTs [20] based on the SDT was used as an analytical matrix in the deductive stage of the qualitative content analysis in this study. In 2019, Gillison et al published a meta-analysis of SDT-based techniques to support health behavior change [19]. This meta-analysis presents an overview of techniques used in SDT interventions and could have been used as an analytical matrix. However, we found the techniques in the classification of MBCTs to be more comprehensively described and therefore easier to use in the analysis, with the additional advantage of being derived from an expert consensus.

In our study, participants expressed support for behavior change in the Safe Step program related to 11 of 21 MBCTs. However, unrepresented MBCTs may still be present in the program but were not specifically asked about in the data collection and were therefore not expressed (eg, the noncontrolling and positive language used in the messages sent by the virtual physical therapist). The SDT has been found applicable to the context of physical activity [35], older adult exercise behaviors [40], and eHealth [42]. However, in relation to our fully self-managed digital context, we encountered a few difficulties when using the classification system. To overcome this, we had to interpret some MBCTs to the fully self-managed and digital context of our study, as many of the MBCTs are facilitated by the involvement of a health care professional or social agent. In order to use the classification of MBCTs to enable better comparisons among self-determination–based interventions in the health domain, especially in this digital era, we suggest that the definitions of the MBCTs should be broadened and perhaps also clarified. For example, the ability to see the exercises in video format and the ease of use of the digital exercise program were in our study stated as important to build confidence in performance and support independence. We were unable to classify these techniques used to support actual performance and increase confidence in performing health-related behaviors. Of note, the analysis contributed with a proposal for an additional competence-based MBCT addressing supportive structures for the performance of behavior.

Methodological trustworthiness has been strived for in various ways in the execution and writing of this study [22]. Three data collection methods were used, including self-reported questionnaires, and individual and focus group interviews. The triangulation of data collection methods enabled looking at data from different perspectives. With the addition of individual interviews and comments in the questionnaires, all participants using the Safe Step program in the feasibility study were represented in the data collection. Overall, older adults with different ages, technology experiences, and exercise adherences were represented in both the individual and focus group interviews to capture the diversity of experiences [43]. Unfortunately, in this study, group compositions regarding women and men became uneven in the focus group interviews, which may have influenced the responses. Moreover, we recognize the fact that it would have been valuable to explore the views of participants who discontinued the intervention in order to explore possible frustration of the basic psychological needs. Furthermore, the experiences of those conducting the interviews varied, which resulted in both novice and in-depth follow-up questions. The analysis was continuously discussed and triangulated by the authors who had different methodological and theoretical backgrounds, professional experiences, or insider and outsider perspectives, which strengthened the credibility of the results. To enable assessment of transferability to other study contexts and target groups/populations, we have described the participants, research context, and methods of the study thoroughly, according to recommendations on how to improve transferability in qualitative studies [23]. We have also strived for a rich description of participants' experiences, with exemplifying quotes.

Conclusions

This study demonstrated that older adults using a fully self-managed digital exercise program for fall prevention (the Safe Step program) expressed support for all three basic psychological needs, though autonomy and competence were predominantly supported compared with relatedness. The Safe Step program supported the development of new exercise routines, and the program was found adaptable to one's own capacities and objectives, and therefore fostered feelings of engagement and ownership in the self-management of exercise. Self-managing fall prevention exercise was found to entail both feelings of freedom and restriction. By using the classification system for MBCTs as an analytical matrix, suggestions for further development of the classification system were made to better suit more digital health interventions. This study proposes to add an MBCT to the classification system that captures support to strengthen a person's competence in performing an activity or task.

XSL•FO
RenderX

Acknowledgments

We would like to thank all older adults who participated in the development of the Safe Step program, and the older adults who participated in the Safe Step trial and the interviews. This work was financially supported by the Swedish Research Council (grant number 2015-03481); the Strategic Research Programme in Care Sciences, Umeå University and the Karolinska Institute, Sweden; the Swedish Research Council for Health, Working Life and Welfare (FORTE); and King Gustav V's and Queen Victoria's Freemasons' Foundation.

Authors' Contributions

MS, LLO, BP, and RJ conceptualized the study. MS and LLO funded the study. LLO performed individual interviews. BP and RJ performed the focus group interviews. BP performed the initial analysis of the qualitative data and continually discussed the analysis with all authors. BP was mainly responsible for drafting the manuscript, with contributions from MW, LLO, and MS. All authors revised and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of the Safe Step project according to the complex intervention framework.
[PNG File , 989 KB - jmir_v23i7e26235_app1.png]

Multimedia Appendix 2
Overview of the areas of questions in the different data collections.
[PNG File , 622 KB - jmir_v23i7e26235_app2.png]

References

1. Hopewell S, Adedire O, Copsey B, Boniface GJ, Sherrington C, Clemson L, et al. Multifactorial and multiple component interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2018 Jul 23;7:CD012221 [FREE Full text] [doi: 10.1002/14651858.CD012221.pub2] [Medline: 30035305]
2. Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. Cochrane Database Syst Rev 2019 Jan 31;1:CD012424 [FREE Full text] [doi: 10.1002/14651858.CD012424.pub2] [Medline: 30703272]
3. Nyman S, Victor C. Older people's participation in and engagement with falls prevention interventions in community settings: an augment to the Cochrane systematic review. Age Ageing 2012 Jan;41(1):16-23. [doi: 10.1093/ageing/afr103] [Medline: 21875865]
4. Hawley-Hague H, Horne M, Campbell M, Demack S, Skelton DA, Todd C. Multiple levels of influence on older adults' attendance and adherence to community exercise classes. Gerontologist 2014 Aug 30;54(4):599-610 [FREE Full text] [doi: 10.1093/geront/gnt075] [Medline: 23899623]
5. Room J, Hannink E, Dawes H, Barker K. What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review. BMJ Open 2017 Dec 14;7(12):e019221 [FREE Full text] [doi: 10.1136/bmjopen-2017-019221] [Medline: 29247111]
6. Gardner B, Jovicic A, Belk C, Kharicha K, Iliffe S, Manthorpe J, et al. Specifying the content of home-based health behaviour change interventions for older people with frailty or at risk of frailty: an exploratory systematic review. BMJ Open 2017 Feb 09;7(2):e014127 [FREE Full text] [doi: 10.1136/bmjopen-2016-014127] [Medline: 28183809]
7. Peters D, Calvo RA, Ryan RM. Designing for Motivation, Engagement and Wellbeing in Digital Experience. Front Psychol 2018 May;9:797-719 [FREE Full text] [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
8. Delbaere K, Valenzuela T, Woodbury A, Davies T, Yeong J, Steffens D, et al. Evaluating the effectiveness of a home-based exercise programme delivered through a tablet computer for preventing falls in older community-dwelling people over 2 years: study protocol for the Standing Tall randomised controlled trial. BMJ Open 2015 Oct 22;5(10):e009173 [FREE Full text] [doi: 10.1136/bmjopen-2015-009173] [Medline: 26493461]
9. van Het Reve E, Silveira P, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: part 2 of a phase II preclinical exploratory trial. J Med Internet Res 2014 Jun 25;16(6):e159 [FREE Full text] [doi: 10.2196/jmir.3055] [Medline: 24966165]
10. Schwenk M, Bergquist R, Boulton E, Van Ancum JM, Nerz C, Weber M, et al. The Adapted Lifestyle-Integrated Functional Exercise Program for Preventing Functional Decline in Young Seniors: Development and Initial Evaluation. Gerontology 2019 May 21;65(4):362-374 [FREE Full text] [doi: 10.1159/000499962] [Medline: 31112941]

XSL•FO
RenderX

11. Pettersson B, Wiklund M, Janols R, Lindgren H, Lundin-Olsson L, Skelton DA, et al. 'Managing pieces of a personal puzzle' - Older people's experiences of self-management falls prevention exercise guided by a digital program or a booklet. BMC Geriatr 2019 Feb 18;19(1):43 [FREE Full text] [doi: 10.1186/s12877-019-1063-9] [Medline: 30777026]
12. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Int J Med Inform 2014 Jun;83(6):416-426. [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]
13. Irvine AB, Gelatt VA, Seeley JR, Macfarlane P, Gau JM. Web-based intervention to promote physical activity by sedentary older adults: randomized controlled trial. J Med Internet Res 2013 Feb 05;15(2):e19 [FREE Full text] [doi: 10.2196/jmir.2158] [Medline: 23470322]
14. Silveira P, van de Langenberg R, van Het Reve E, Daniel F, Casati F, de Bruin ED. Tablet-based strength-balance training to motivate and improve adherence to exercise in independently living older people: a phase II preclinical exploratory trial. J Med Internet Res 2013 Aug 12;15(8):e159 [FREE Full text] [doi: 10.2196/jmir.2579] [Medline: 23939401]
15. French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct 20;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
16. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Press; 2018.
17. Ryan RM. Psychological needs and the facilitation of integrative processes. J Pers 1995 Sep;63(3):397-427. [doi: 10.1111/j.1467-6494.1995.tb00501.x] [Medline: 7562360]
18. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
19. Gillison FB, Rouse P, Standage M, Sebire SJ, Ryan RM. A meta-analysis of techniques to promote motivation for health behaviour change from a self-determination theory perspective. Health Psychol Rev 2019 Mar 16;13(1):110-130. [doi: 10.1080/17437199.2018.1534071] [Medline: 30295176]
20. Teixeira PJ, Marques MM, Silva MN, Brunet J, Duda JL, Haerens L, et al. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. Motivation Science 2020 Dec 05;6(4):438-455. [doi: 10.1037/mot0000172]
21. Mansson L, Lundin-Olsson L, Skelton DA, Janols R, Lindgren H, Rosendahl E, et al. Older adults' preferences for, adherence to and experiences of two self-management falls prevention home exercise programmes: a comparison between a digital programme and a paper booklet. BMC Geriatr 2020 Jun 15;20(1):209 [FREE Full text] [doi: 10.1186/s12877-020-01592-x] [Medline: 32539711]
22. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
24. Ghaye T, Melander-Wikman A, Kisare M, Chambers P, Bergmark U, Kostenius C, et al. Participatory and appreciative action and reflection (PAAR) - democratizing reflective practices. CREP 2008 Nov;9(4):361-397. [doi: 10.1080/14623940802475827]
25. Sandlund M, Lindgren H, Pohl P, Melander-Wikman A, Bergvall-Kåreborn B, Lundin-Olsson L. Towards a mobile exercise application to prevent falls: a participatory design process. Int J Child Health Hum Dev 2016;9(3):389-398.
26. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
27. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
28. Lindgren B, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud 2020 Aug;108:103632. [doi: 10.1016/j.ijnurstu.2020.103632] [Medline: 32505813]
29. Patall EA, Cooper H, Robinson JC. The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. Psychol Bull 2008 Mar;134(2):270-300. [doi: 10.1037/0033-2909.134.2.270] [Medline: 18298272]
30. Ehn M, Johansson A, Revenäs Å. Technology-Based Motivation Support for Seniors' Physical Activity-A Qualitative Study on Seniors' and Health Care Professionals' Views. Int J Environ Res Public Health 2019 Jul 08;16(13):2418 [FREE Full text] [doi: 10.3390/ijerph16132418] [Medline: 31288398]
31. Finnegan S, Bruce J, Seers K. What enables older people to continue with their falls prevention exercises? A qualitative systematic review. BMJ Open 2019 Apr 15;9(4):e026074 [FREE Full text] [doi: 10.1136/bmjopen-2018-026074] [Medline: 30992291]
32. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

XSL•FO
RenderX

33. Aelterman N, Vansteenkiste M, Haerens L, Soenens B, Fontaine JRJ, Reeve J. Toward an integrative and fine-grained insight in motivating and demotivating teaching styles: The merits of a circumplex approach. Journal of Educational Psychology 2019 Apr;111(3):497-521. [doi: 10.1037/edu0000293]
34. Schwaninger I, Frauenberger C, Fitzpatrick G. Unpacking Forms of Relatedness around Older People and Telecare. In: DIS' 20 Companion: Companion Publication of the 2020 ACM Designing Interactive Systems Conference. 2020 Jul Presented at: 2020 ACM Designing Interactive Systems Conference; July 6-10, 2020; Eindhoven, Netherlands p. 163-169. [doi: 10.1145/3393914.3395867]
35. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
36. Markland D, Tobin VJ. Need support and behavioural regulations for exercise among exercise referral scheme clients: The mediating role of psychological need satisfaction. Psychology of Sport and Exercise 2010 Mar;11(2):91-99. [doi: 10.1016/j.psychsport.2009.07.001]
37. Barmentloo LM, Olij BF, Erasmus V, Smilde D, Schoon Y, Polinder S. Personal preferences of participation in fall prevention programmes: a descriptive study. BMC Geriatr 2020 May 28;20(1):185 [FREE Full text] [doi: 10.1186/s12877-020-01586-9] [Medline: 32466747]
38. Breda AI, Watts AS. Expectations Regarding Aging, Physical Activity, and Physical Function in Older Adults. Gerontol Geriatr Med 2017 Apr 11;3:2333721417702350 [FREE Full text] [doi: 10.1177/2333721417702350] [Medline: 28491915]
39. Massie AS, Meisner BA. Perceptions of aging and experiences of ageism as constraining factors of moderate to vigorous leisure-time physical activity in later life. Loisir et Société / Society and Leisure 2019 Mar 12;42(1):24-42. [doi: 10.1080/07053436.2019.1582903]
40. Kirkland RA, Karlin NJ, Stellino MB, Pulos S. Basic Psychological Needs Satisfaction, Motivation, and Exercise in Older Adults. Activities, Adaptation & Aging 2011 Jul;35(3):181-196. [doi: 10.1080/01924788.2011.596764]
41. Sandlund M, Pohl P, Ahlgren C, Skelton DA, Melander-Wikman A, Bergvall-Kåreborn B, et al. Gender Perspective on Older People's Exercise Preferences and Motivators in the Context of Falls Prevention: A Qualitative Study. Biomed Res Int 2018;2018:6865156 [FREE Full text] [doi: 10.1155/2018/6865156] [Medline: 30112416]
42. Pingree S, Hawkins R, Baker T, duBenske L, Roberts LJ, Gustafson DH. The value of theory for enhancing and understanding e-health interventions. Am J Prev Med 2010 Jan;38(1):103-109 [FREE Full text] [doi: 10.1016/j.amepre.2009.09.035] [Medline: 20117565]
43. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]

Abbreviations

MBCT: motivation and behavior change technique
SDT: self-determination theory

Edited by R Kukafka; submitted 11.12.20; peer-reviewed by P Anderberg, K Fitzner, M Ehn; comments to author 27.02.21; revised version received 24.03.21; accepted 30.04.21; published 30.07.21.

Please cite as:
Pettersson B, Janols R, Wiklund M, Lundin-Olsson L, Sandlund M
Older Adults' Experiences of Behavior Change Support in a Digital Fall Prevention Exercise Program: Qualitative Study Framed by the Self-determination Theory
J Med Internet Res 2021;23(7):e26235
URL: https://www.jmir.org/2021/7/e26235
doi:10.2196/26235
PMID:34328438

©Beatrice Pettersson, Rebecka Janols, Maria Wiklund, Lillemor Lundin-Olsson, Marlene Sandlund. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Supporting Patient-Clinician Interaction in Chronic HIV Care: Design and Development of a Patient-Reported Outcomes Software Application

Susan Herrmann[1], PhD; Brad Power[2], BSc; Amineh Rashidi[3], PhD; Mark Cypher[2], PhD; Frank Mastaglia[2], BCom; Amy Grace[4], BPsych; Elizabeth McKinnon[5], PhD; Pierre Sarrot[6], MSc; Christophe Michau[7], MD; Matthew Skinner[8], MBBS, FRACP; Renae Desai[9], MBBS, MD, FRCP, FRACP; Martin Duracinsky[6,10], MD, PhD

[1]Medical School, University of Western Australia, Crawley, Australia

[2]College of Arts, Business, Law & Social Sciences, Murdoch University, Murdoch, WA, Australia

[3]School of Nursing & Midwifery, Edith Cowan University, Joondalup, WA, Australia

[4]School of Psychology and Exercise Science, Murdoch University, Murdoch, WA, Australia

[5]Telethon Kids Institute, Nedlands, WA, Australia

[6]Patient-Centered Outcomes Research, Health Economics Clinical Trial Unit, Hospital Hotel-Dieu, AP-HP, University de Paris, Paris, France

[7]CeGIDD, Centre Hospitalier Saint Nazaire, Saint Nazaire, France

[8]Sir Charles Gairdner Hospital, Nedlands, WA, Australia

[9]Perron Institute for Neurological & Translational Science, Queen Elizabeth II Medical Centre, Nedlands, WA, Australia

[10]Department of Internal Medicine & Immunology, Hospital Bicetre, AP-HP, Kremlin-Bicetre, France

Corresponding Author:
Susan Herrmann, PhD
Medical School
University of Western Australia
35 Stirling Hwy
Crawley, 6009
Australia
Phone: 61 0417997934
Email: Susan.Herrmann@uwa.edu.au

Abstract

Background: The consideration of health-related quality of life (HRQL) is a hallmark of best practice in HIV care. Information technology offers an opportunity to more closely engage patients with chronic HIV infection in their long-term management and support a focus on HRQL. However, the implementation of patient-reported outcome (PRO) measures, such as HRQL in routine care, is challenged by the need to synthesize data generated by questionnaires, the complexity of collecting data between patient visits, and the integration of results into clinical decision-making processes.

Objective: Our aim is to design and pilot-test a multimedia software platform to overcome these challenges and provide a vehicle to increase focus on HRQL issues in HIV management.

Methods: A multidisciplinary team in France and Australia conducted the study with 120 patients and 16 doctors contributing to the design and development of the software. We used agile development principles, user-centered design, and qualitative research methods to develop and pilot the software platform. We developed a prototype application to determine the acceptability of the software and piloted the final version with 41 Australian and 19 French residents using 2 validated electronic questionnaires, the Depression, Anxiety and Stress Scale-21 Items, and the Patient Reported Outcomes Quality of Life-HIV.

Results: Testing of the prototype demonstrated that patients wanted an application that was intuitive and without excessive instruction, so it felt effortless to use, as well as secure and discreet. Clinicians wanted the PRO data synthesized, presented clearly and succinctly, and clinically actionable. Safety concerns for patients and clinicians included confidentiality, and the potential for breakdown in communication if insufficient user training was not provided. The final product, piloted with patients from both countries, showed that most respondents found the application easy to use and comprehend. The usability testing survey administered found that older Australians had reduced scores for understanding the visual interface (P=.004) and finding the *buttons* organized (P=.02). Three-fourths of the respondents were concerned with confidentiality (P=.007), and this result was more prevalent in

XSL•FO
RenderX

participants with higher anxiety and stress scores (P=.01), as measured by the Depression, Anxiety and Stress Scale-21 Items. These statistical associations were not observed in 15 French patients who completed the same questionnaire.

Conclusions: Digital applications in health care should be safe and fit for purpose. Our software was acceptable to patients and shows potential to overcome some barriers to the implementation of PROs in routine care. The design of the clinicians' interface presents a solution to the problem of voluminous data, both synthesizing and providing a snapshot of longitudinal data. The next stage is to conduct a randomized controlled trial to determine whether patients experience increased satisfaction with care and whether doctors perceive that they deliver better clinical care without compromising efficiency.

(J Med Internet Res 2021;23(7):e27861) doi:10.2196/27861

KEYWORDS

physician-patient communication; eHealth; primary care; ambulatory care; information systems; user-centered design; user context; patient reported outcomes; qualitative research; health care; HIV care; mobile phone

Introduction

Background

Chronic HIV infection is a complex disease associated with psychosocial morbidity that can affect health-related quality of life (HRQL) detrimentally [1]; however, as clinical management in primary care settings often focuses on physical well-being and the outcomes of antiretroviral therapy [1], psychosocial and other relevant patient-volunteered information may not be engaged in the same clinically systematic manner. Challenges to implementing patient-reported outcome (PRO) measurements in routine care include the synthesis of voluminous data generated by questionnaires regardless of whether they are e-questionnaires or paper based, the complexity of collecting data between patient visits, and the integration of the results into clinical decision-making protocols [2]. There are additional organizational barriers to integrating PRO measurement in real-life settings, in hospitals and at home, but guidance is available [3]; and obstacles have also been recognized in localized settings [4]. As some barriers become surmountable [2], there is an increased impetus in fields, such as HIV and oncology, to use validated measures for the assessment of physical and psychosocial symptoms and quality of life beyond the hospital setting and between patient visits [5-7]. Muessig et al [7] identified opportunities for digital health strategies to be integrated within the HIV Continuum of Care to aid retention. Those concerned with cancer care and trials have taken advantage of the advances in digital communications [8] to configure hospital computer systems and capture PRO data in real time to better reflect the patient experience [9] and facilitate intervention. Kjaer et al [10] described the hospital-based implementation of a web-based tool to facilitate the use of PROs in HIV care. They adapted AmbuFlex [11], a generic web-based program for the collection and synthesis of PRO data, envisaging that this platform could individualize patient care and inform targeted resource allocation, particularly with regard to the frequency of patient medical appointments. In the context of the current COVID-19 pandemic, reducing face-to-face encounters without compromising clinical outcomes is of current interest to both patients and health professionals [12]. However, although digital innovation may offer solutions to problems, new systems must be accepted and regarded as advantageous to patients and worthwhile for providers to implement successfully in clinical practice [13].

Objectives

We aim to develop a digital platform to facilitate the use of PROs in the context of chronic HIV infection. We believe that the meaningful engagement of health providers and patients, as equal stakeholders, will create a bidirectional incentive to use PRO software whereby "a common patient-centered frame of reference" [3] supports therapeutic communication and patient autonomy. Therefore, the digital communication package that we designed should reflect stakeholders' collective and differing requirements. We opted to use an iterative development process of agile design and qualitative research methods [14-16], in contrast with earlier approaches to software design that, in targeting potential users, tended to focus on technological capacity and maximum specification.

We envisaged the software to support the following communication process: patients would input information at home or in the clinic before their appointments using their own devices. The software then transmits the information to the medical practice software through a secure portal. Using a custom interface, doctors could review concise data from an automated, preprocessed synthesis of patient self-reports during patient consultation. The longitudinal information displayed graphically provides clinicians with trends over time. We set the following objectives: (1) to design a web-based application compatible with mobile and desktop, (2) to assess comprehension of the platform, and perceived benefits, (3) to assess the usability of the designed product, and (4) to conduct a pilot study using PRO instruments integrated into the platform. Our interdisciplinary Australian and French teams included medical, psychology, information technology (IT), and digital communications experts, and the study was conducted in both countries with the teams holding regular teleconferences over the course of the study.

Methods

Study Design

The study was observational and iterative, with each activity informing the next. In stage 1, we created and tested a demonstration prototype or *wireframe* and assessed patient and clinician users' comprehension of the rationale behind the software. We then refined the design further, loaded the software with two validated PRO instruments, the Depression, Anxiety and Stress Scale (DASS-21) [17] and the Patient Reported

XSL•FO
RenderX

Outcomes Quality of Life-HIV (PROQOL-HIV) HRQL instrument specific for HIV [18]; in stage 2, we initiated a pilot study. The project was conducted between December 2015 and May 2017, and the Murdoch University Human Research Ethics Committee (2015/228) and the French Ethics Committee Ile de France IV–Institutional Review Board 00003835 (2016/44NI) approved the study. The research was conducted in 3 hospitals in France and 2 specialist HIV practices in Western Australia (WA). Members of the research team with fluency in both languages carried out translations when necessary.

Participants

Participants were a sample of convenience and were eligible for the study if they were at least 18 years old and able to provide informed consent. All patients had been diagnosed with HIV infection and presented with a wide range of clinical and sociodemographic characteristics. Clinicians were trained either in infectious diseases or immunology, or were general practitioners with training in HIV medicine. Participants in stage 1 included patients with chronic HIV infection (Australian=4, French=36) and clinicians (Australian=4, French=12); those in stage 2 were patients (Australian=41, French=39) recruited from 3 hospitals in Paris and 2 community-based HIV specialist centers in WA. In addition, colleagues with various professional backgrounds, including clinicians, from within the 2 research organizations were asked to carry out the initial user testing of a basic prototype (stage 1).

Data Analysis

An interview guide was developed collaboratively. One-on-one face-to-face interviews, up to 45 minutes in length, were conducted in the study hospitals in France and at specialist practices in Australia. The interviews were recorded and transcribed. The transcripts were imported to NVivo 12 (QSR International Pty Ltd.) and analyzed independently by a male master's student, a female PhD student, and their PhD supervisors (SH, MD). An inductive thematic approach [19], informed by field data, was used to analyze the interview data. We coded text from the interviews to common themes and continued until the data were considered saturated [20]. Questionnaire data from the pilot study were exported and encrypted from the web application to the data capture program REDCap (Research Electronic Data Capture) [21] and subsequently downloaded into the software package R for quantitative analysis [22]. This included the derivation of descriptive summaries and the application of linear or logistic regression to explore associations with aspects of usability. Demographic information is presented as percentages or means, as appropriate.

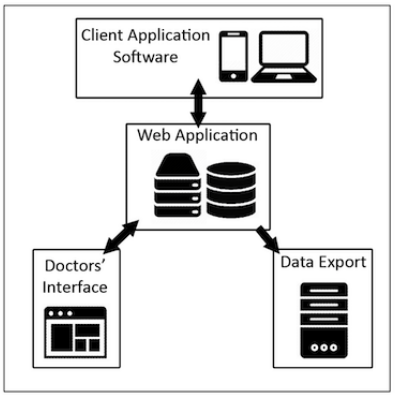
Stage 1: Design and Testing of the Initial Prototype

To approach the design of the software, we implemented a user-centered design [23] and agile development principles [14,15] that promote adaptive planning, evolutionary development, early delivery, and continuous improvement. This methodology provided increased transparency during the development phase through continuous feedback and flexibility when changes in direction were required. Our agile approach was informed by narratives, and as such, we educated the team on the lived patient experience of people affected by HIV by writing fictional personas informed by previous qualitative research [1].

The prototype consisted of basic functions, representing a minimum viable product. The architecture (Figure 1) comprises a web application that allowed qualitative instruments to be configured and patient accounts to be created. The instruments can then be deployed to client software applications that run on either mobile or desktop platforms and allow patients to provide responses to questionnaires. The client then submits the patient's response data via a secure channel back to the web application, where it is stored in a database. The doctors' interface was accessed through a web browser. This allows doctors and clinicians to access processed graphical reports of the scored patient responses to a PRO, identify longitudinal trends, and promote discussion with the patient. Finally, data export functionality allows the web application to export data in suitable formats for import into clinical patient management software. This architecture was implemented as a first pass to provide a minimum viable product for the initial testing.

We invited 20 colleagues, including clinicians, from both teams to use and comment on the software, which was configured with 10 testing items comprising questions and response categories. A member of the research team observed and recorded the users' verbal responses as they navigated the application and administered a short questionnaire. The data were collated as *first round, in-house user testing* to improve the usability of the prototype.

Figure 1. Diagram of communication between platforms.



XSL•FO
RenderX

Acceptability of the Concept

Contemporaneously, we conducted 12 semidirected interviews of French patients to establish patients' perceptions on using health care–related software and the difference eHealth might make to their care. Other questions were related to the current style of communication between patients and clinicians and their experiences with electronic devices and mobile health (mHealth) apps.

Testing of the Prototype

The second round of user testing, comprising 44 semidirected interviews, took place in 3 Paris hospitals and 1 community-based clinic in WA. An interview guide was developed. In the first part of the interview, researchers described the purpose of the application and asked the patients (n=28) and doctors (n=16) about their computer usage habits and capabilities and their views on perceived benefits or harms of the software. They were then given an opportunity to view the application, and their reactions were noted [24]. Through this before-and-after approach, we were able to identify discrepancies between participants' expectations of the software and subsequent user experience.

Graphic Design

Informed by the fictional personae and predesign interviews, *mood boards* were created for colors, icons, and text (Multimedia Appendix 1). The initial design was for mobile devices such as smartphones, with the small screen presenting the greater design challenge regarding functionality. However, on a larger screen, there was no loss of function or esthetics (Figure 2). We used the composite data, empirical observations, previous research, and preliminary in-house user testing to create *user stories*, with short statements representing user requirements in French and English, which justified new features (Textbox 1). Each user story was coded to track the origin of the story from patient interview or empirical data (eg, "As a <stakeholder type> I want <some feature> so that <some benefit>"). Expressed in this way, it made the features easier to talk about and remember who they were for and to prioritize feature development. The stakeholders were patients, clinicians, researchers, and developers. The list of user stories informed a user story "backlog"—like a waiting room, ready to inform sprints that denoted a period of software development.

Figure 2. App design.



Textbox 1. User stories and sample quotes.

Sample Quotes

- "As a patient, I want to be in control of my data, and know it is secure and private, so that I can feel comfortable providing it."
- "As a patient, I want this software available on the types of device(s) I own, so that it is accessible to me."
- "As a patient, I want clean, distraction free interaction with appropriate detail for fine control, so that my data are precise."
- "As a patient, I want to use the software intuitively, without excessive instruction, so that it feels easy and effortless."
- "As a patient, I want a personalized experience, so that it does not feel like filling out a form."
- "As a doctor, I want clear and succinct results that inform a consultation, so that I do not get overwhelmed."
- "As a doctor, I want unbiased results from the digital instruments, so that the patient reported outcomes are accurate."

Design of the Clinician Interface

In Australia, we interviewed 4 doctors, comprising 2 men and 2 women; 3 were from HIV specialist practices and 1 was hospital-based. We interviewed them before and after showing them the clinician interface (wireframe). The doctors were asked to envisage the utility of the software for measurement of HRQL and, informed by this feedback, the graphic artist refined the design of the interface (Figure 2).

XSL•FO
RenderX

Stage 2: Pilot Studies in Australia and France

Australia

We recruited 41 Australians who were undergoing treatment for HIV infection from 2 high-case load community practices by the research nurse. After receiving minimal instruction and choosing a device (ie, desktop, laptop, tablet, or smartphone), they created an account and accessed the platform. Several participants preferred to use the software in their homes after phone contact with the researcher, but most accessed it in the clinic. The platform was configured using 4 questionnaires, including the 2 validated instruments: DASS-21 [17] and PROQOL-HIV [18,25]. The results of the 2 PROs are reported in this manuscript. Participants verbalized their experience while navigating the software to enable the researcher to note their observations, and they were instructed to record *true responses*. Participants also completed a usability questionnaire, and the researcher ranked the frequency of repetitive (similar) comments and provided a report to the developer.

France

The implementation of the pilot in France was similar to that in Australia. Participants included 39 outpatients from 2 centers, comprising 20 from the Paris area hospital, Kremlin Bicetre, and 19 in Saint Nazare, a regional hospital. Before the consultation, a research assistant introduced the study, and patients completed the questionnaires using a tablet or computer before consulting the doctor. Comments from this observation process and from participants were noted and synthesized.

Results

Stage 1: Design and Testing of the Initial Prototype

Sociodemographic Data

The sociodemographic background of the patient participants in France and Australia included in the study were different. The participants in Australia were predominantly White men, whereas 42% (10/24) of the participants in France were women, and in general, were more culturally and linguistically diverse, with 67% (16/24) from the African continent. All participants completed at least secondary level education.

Acceptability of the Concept: Patients' Perspectives of Using Digital Health Technology

French Participants

Semidirected interviews were conducted with 24 patients in France, of which 92% (22/24) owned a computer, a third used a tablet, and half possessed smartphones. These interviews included 12 participants who had not seen the application and an additional 12 who reviewed the application after a demonstration using the wireframes to explain the features, by the researcher.

The use of mHealth and digital information sources among these respondents consisted mainly of fitness trackers, government services, and HIV information or news sites. Discussions included the concept of using mHealth devices to communicate HRQL issues with their doctors. Although 75% (18/24) of patients foresaw advantages to the application as described to them, the other 25% (6/24) could not see the point of using technology in place of face-to-face communication. Some respondents liked the idea of being able to contact their doctors directly, especially in an *emergency*, and 17% (4/24) did this through email. Others saw opportunities for the application to act as a reminder, thus focusing the consultation on the topics important to them. For example, one patient stated, "Sometimes when we meet, we do not remember what we wanted to say." Another respondent thought it would strengthen engagement between doctor and patient:

> *In the health domain it will be a good tool...will bring patients closer to their doctors, it's great...*

However, for some participants, an mHealth evaluation of *quality of life* was not of particular value. One participant stated the following:

> *I do not see any utility, I am already followed, I have already been HIV positive for several years and I do the things (sic) well, it is well controlled. I have an appointment every six months. During the six months I do not see any usefulness of an application.*

Some patients thought that the volume of data generated would overload their doctors and reduce the consultation time available for personal conversation. With regard to confidentiality, one patient described himself as "suspicious" of "these types of things." An overriding concern for 45% (11/24) of French participants, particularly those with migrant backgrounds, was data security and confidentiality. One participant stated as follows:

> *You do not know to whom you can share all your life, you spread all your life like that; after you do not know who can recover this..., all that is not necessarily good.*

The concern was that their information would be used in newspapers, and they would be identified and exposed to stigma and discrimination, or that information would be revealed if devices were lost. Our conclusion from this round of testing was that patients' acceptance of a potential digital health care communication pathway in France was low.

Australian Participants

In Australia, 4 in-depth interviews were conducted. The comments of the participants before being shown the prototype application were similar to those of the French participants but reflected more optimism about eHealth services and less concern about confidentiality. In contrast to French participants, Australian participants used the prototype application and progressed through a series of questions drawn from the HRQL questionnaire. The screen design was minimal, showing one question per screen, and was accompanied by a *peaceful* soundtrack. The participants' feedback after using the software was largely positive, and participants envisaged that doctors would have a better picture of their patients' current health state if the information derived from the HRQL instrument was actively engaged with during the consultation; here, the caveat was how the general practitioner interpreted the patient information. There was a perception that digital communication in health was inevitable with one patient stating as follows:

XSL•FO
RenderX

> *I think it is good...The whole world works on technology, so the medical system may as well.*

Another patient foresaw benefits to discourse within a prescribed timeframe: "It would be good for the software to be able to let me answer at my home at my own pace and [information] would be shared with my doctor." However, one patient commented as follows:

> *OK, so you can get a prompt current reply with the technology, you don't have to make an appointment to see a doctor, you can get it via technology.*

The patient perceived this as a disadvantage that would potentially complicate his care, but the comment also suggested that the patient envisaged an electronic response from the doctor, rather than unidirectional communication, which was used to enhance face-to-face communication. Another patient considered his future clinical care:

> *I think for me, because I'm young and I don't have a lot of effects from it [HIV] yet. I guess as I get older, I'm probably going to have more effects, so it's able to keep a really good record of the effects on me, and the doctors always having that record, because it's technology-based, it's never going to go missing.*

In contrast to French participants, concern about confidentiality among Australian participants was low. However, one participant would not complete the sensitive information on a portable device when the screen was visible to others. Vulnerability to hacking was also a concern, with a participant stating as follows:

> *It's not so much you using the data or me giving the data to you...but if you're ever hacked.*

However, another participant said as follows:

> *No, I am not worried, I think confidentiality, like all software, is the same as on paper. You know if anyone wants to get into your documentation they need your passwords or something like that, so, it's capable of doing that nowadays, if you're doing banking and everything...*

Design and Usability

The Australian group provided further detailed feedback about the design and usability of the software, commenting on the mouse versus touch screen and the navigation icons and their likes and dislikes. Consequently, the developer improved the product, removed the soundtrack, and added a privacy statement concerning the level of data encryption provided in the application.

Clinicians' Perspectives: France and Australia

We deduced from our interviews with clinicians (12 in France and 4 in Australia) that three key design features were important. First, the PRO instrument chosen should be valid and disease specific. Second, it could be used to record and monitor outcomes of interventions and the impacts of the disease process. Finally, the application should be frictionless and integrated with the flow of communication between patient and clinician, as well as with other programs used simultaneously, which may require *screen switching*.

Younger clinicians had a more positive view of the potential use of mHealth applications than their older colleagues, and they valued digital clinical support software (eg, for prescribing medications). One doctor actively promoted mHealth apps to his patients for self-monitoring and motivational feedback. They also expressed interest in programs to increase retention and continuity of care, thus agreeing with the patient participants' views on the application. One doctor said the following:

> *The benefit is certain...Because HIV patients are particular, [HIV] always has an impact on their lives, whether it is social or quality of life.*

The 4 Australian doctors envisaged how it might be useful for chronic HIV management, as managed by nurses. They recognized the potential for integrating features to support the monitoring of treatment adherence and healthy lifestyle assessments (eg, drug and alcohol use).

Overwhelmed With Information

In general, doctors were concerned about information overload, with numerous complex concerns becoming difficult to address in the consultation time available. One doctor said, "...the data (for PROs) may be too complex for the average practitioner to interpret" and that, "an amateur assessment of exercise, for example, may not be better than no assessment of exercise."

This clinician articulated that "doctors feel like they need to fix things," implying that PROs might present them with information that could not be adequately actioned, leading to an unsatisfactory outcome for the patient. Another doctor said as follows:

> *It [the software] has to provide me with a reasonable summary of what is going on without taking 15 minutes to digest it. Otherwise, it just gets too much, does it provide value?...and people will not get through their work, and they will just ignore it.*

French hospital-based doctors stressed that the time available within their consultations to review HRQL via a software application was extremely limited. In primary care, doctors tend to work within purpose-built IT systems for general practice. These programs often contain short questionnaires, such as an alcohol assessment; however, in hospitals, there is a pressing need for an integrated system of IT programs. One doctor stated as follows:

> *...there are a lot of applications that are not integrated...and you have to open up 5 different screens to access 5 different aspects of their (patients) care.*

After seeing a representation of the doctors' interface, the 3 primary care clinicians noted the utility of having patient-reported HRQL synthesized and particular items flagged for attention, thus saving time and facilitating what one doctor described as "a better structure to the consultation." One doctor noted that the software could benefit patients with memory problems.

XSL•FO
RenderX

Patient Safety

One of the hospital-based clinicians was less convinced of the value of the software for individual patient care because patients' queries are triaged by nurses, whereas in the primary care setting, communication is largely within the consultation and triaged by the doctor. Consequently, one doctor said, "it is very important for the patients not to consider that by pressing 'send' they are directly communicating urgently to the doctor." Clinicians placed emphasis on the need to manage patient expectations and the necessity for clinician training to use and interpret PRO data, so that it can be actioned appropriately.

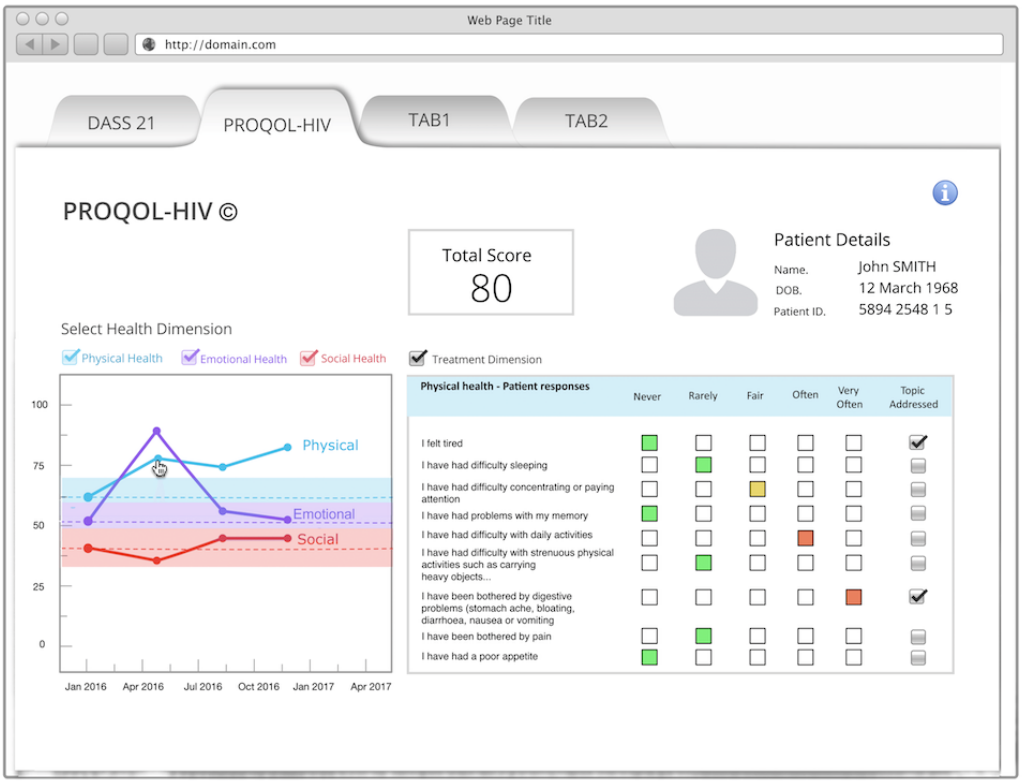
Confidentiality

In France, doctors were more likely to address the issue of data security rather than patient confidentiality. They considered *absolute security* as a critical element of the application. Doctors mentioned several potential risks related to the use of such an application, namely data storage on mobile devices, data submission via the internet, and data security in hospital servers.

Clinicians' Interface

The clinicians' interface (Figure 3) addresses concerns about the volume of granular information generated by PRO instruments. We used the HRQL questionnaire to illustrate the function of the display for PRO data. Briefly, the items displayed on the screen are color-coded to each response (rarely, sometimes, always, etc), with the color red indicating patient concern to emphasize the report as patient-driven. Accordingly, a sidebar checkbox allows the doctor to indicate that a discussion has taken place. Sections pertaining to physical, emotional, and social health dimensions are displayed as graphs to enable the visualization of changes in health dimension scores over time.

Figure 3. Clinicians' interface.



Stage 2: Results of the Pilot Studies in Australia and France

Table 1 shows the characteristics of patient participants in the pilot study. The average age of Australian patients was 57 years and 90% (37/41) were men, and HIV transmission was largely sexual reflecting the demographic of the Australian HIV epidemic [26]. The results of the usability testing (Table 2) showed that 98% (40/41) of the participants found the application easy to use, and 61% (25/41) comprehended it quickly. We also observed that older age was associated with reduced scores for understanding the visual interface (P=.004) and finding the *buttons* organized (P=.02). Furthermore, 24% (10/41) of respondents were concerned with confidentiality, and this result was more prevalent in participants with higher anxiety and stress scores, as measured by the DASS inventory (P=.007 and P=.01, respectively).

The French pilot patient participants' median age was 53 years; they were mostly men and more often single. Usability testing was completed by 19 participants. The results were similar to that of Australian participants regarding ease of use and quick comprehension; however, the French participants were more concerned about confidentiality (9/19, 47%) than the Australian participants. However, for the 15 patients who completed the usability questionnaire and the DASS, the correlation between confidentiality and anxiety and stress scores measured by the DASS inventory was not significant (r=0.37; P=.17). The association between older Australians and concern about the organization and function of the application was not tested in the French participants because most French participants had

XSL•FO
RenderX

assigned a maximum score to those items, creating a *ceiling effect*.

The commentary received from the French patients was largely positive for all aspects of the application design. Overall, patients reported being glad to identify topics that they would not usually or spontaneously discuss with their clinician, and this opportunity could incentivize them to use the application over a longer period. A concern was raised that standardization via the use of questionnaires may threaten individualized care and not take into consideration individual circumstances. For doctors, an annoyance was the necessity to open another software application on their screen, in addition to the patient record.

Table 1. Characteristics of patients in the pilot study (N=80).

Characteristics	Australian patients		French patients	
	Total, n (%)	Value	Total, n (%)	Value
Demographics, median (range)				
Age (years)	41 (51)	58 (50-64)	39 (49)	52 (44-60)
Duration HIV (months)	34 (43)	228 (153-312)	36 (45)	126 (69-240)
Sex, n (%)	41 (51)		36 (45)	
Male		37 (90)		29 (81)
Female		4 (10)		7 (19)
Ethnicity, n (%)	38 (48)		36 (45)	
White		33 (87)		27 (75)
Asian		4 (10)		1 (3)
Middle Eastern		1 (3)		0 (0)
African		0 (0)		7 (19)
Latinx		0 (0)		1 (3)
Mode of transmission, n (%)	33(41)		36 (45)	
MSM[a]		27 (82)		22 (61)
Heterosexual		6 (18)		14 (39)
Employment, n (%)	41 (51)		19 (34)	
Employed		23 (56)		12 (63)
Not employed		18 (44)		7 (37)
Clinical variables				
CD4 T cells, median (range)	31 (39)	703 (458-946)	36 (45)	610 (480-872)
HIV viral load, n (%)				
>20 copies/ml	31 (39)	N/A[b]	33 (41)	33 (92)
<20 copies/ml	31 (39)	31 (100)	3 (4)	3 (8)
HBV-HCV coinfection, n (%)	31 (39)	0 (0)	35 (44)	10 (29)

[a]MSM: men who have sex with men.

[b]N/A: Not applicable.

XSL•FO
RenderX

Table 2. Pilot study usability testing[a] (N=60).

Response[b]	Australian (n=41), n (%)	French (n=19), n (%)
Easy to use	40 (98)	18 (95)
I would need support to use it	0 (0)	3 (16)
I would use it again	41 (100)	16 (84)
Had a clear, clean design	40 (98)	18 (95)
Required minimum screen changes	41 (100)	N/A[c]
I found it easy and quick to comprehend	25 (61)	18 (95)
The instructions were clear and unambiguous	37 (90)	17 (90)
The buttons organized and easy to find	35 (85)	19 (100)
I understood the function	28 (68)	19 (100)
I found it easy to navigate	40 (98)	19 (100)
The size, style, and font were appropriate	39 (95)	N/A
It was a pleasant experience	38 (93)	17 (90)
I found it intuitive to use	31 (76)	15 (79)
I was concerned about confidentiality	10 (24)	9 (47)

[a]Pilot test was carried out in a primary health care setting (Australia) and a hospital-based setting (France).

[b]Responses to the questionnaire item "The application was judged to be."

[c]N/A: Not applicable (two items were not asked of French participants).

Discussion

Principal Findings

We designed and tested a software application enabling the transfer of PRO from a user's device to a clinician's electronic patient management system through a secure portal. A key capability is the display of synthesized data with the potential to focus patient and doctor discourse on patient-driven information; the data are measurable and can be mapped against the implementation of interventions (ie, action points). The design permits the prioritization of actions between visits, further supporting patient preference, and the use of validated instruments could accommodate concerns about the consistency and reliability of repeated PRO results. As the software is configurable, validated PRO instruments can be selected based on clinical requirements (eg, a questionnaire for monitoring side effects or adherence to a new medication). The use of ePRO measures is becoming commonplace and acceptable with some caveats [27]. Our methods enabled us to integrate users' opinions on the application design and increase the validity of the product [28] and its visual appeal. The design is a combination of function and form, as it is esthetically pleasing. However, by taking a minimal approach to the software design, we maximized the potential of the product to meet the needs of the end user and prevent "overspecification." Although we found that older age was associated with reduced scores for some aspects of usability, none of the patients in the pilot study indicated that they would need support to use the application again.

Data from the predesign phase indicated interest in enhanced patient-doctor communication, supported by mHealth strategies, although a few patients considered it unnecessary. Hitherto, most participants had only thought of mHealth in terms of health monitoring and fitness applications. There was initial confusion regarding what form the flow of information would take and what feedback, if any, they would receive from their doctors. The French patients considered feedback important and expected results or a *score*. The Australian patients were more concerned about how the doctor would use this information and whether it would improve their care in the long term. In general, participants envisaged an *efficiency gain* for themselves, and the doctors, if the volume of data transmitted was not overwhelming. The idea of entering data at home in preparation for their medical consult fits with their practice of a preappointment blood test, whereby the results would be ready for discussion on arrival. In both countries, electronic transmission of laboratory results to medical practice software and linkage with the patient's file is common. Confidentiality and data security were major concerns among French patients, particularly migrants, and doctors. Such concerns were less striking among Australian interviewees who had confidence in data protection processes. However, the results of the Australian pilot study revealed that participants with high stress and anxiety scores as measured by the DASS tended to worry about confidentiality when entering their own health information in actuality, as opposed to a hypothetical situation, and this was true of the French participants as well. Evidence also showed that participants perceived greater privacy in using the application on a computer in their own home, in contrast to using the application on a mobile device in public. Notwithstanding, people with stigmatized illnesses are significantly more likely to use the internet to communicate with clinicians [29].

XSL•FO
RenderX

Strengths and Limitations of the Study

We conducted our study in a primary care context in Australia and a hospital-based setting in France. Although this is a strength in the context of product development, it has implications for the design of randomized controlled trials and the subsequent implementation of PRO measures in clinical care [30,31]. We tested our software in the context of HIV care, but we believe our findings are equally useful for other chronic illnesses. Both countries offer free universal health care; however, in Australia, primary care is fee for service and government rebates rarely cover the cost of a consultation, which varies according to length. Public hospital care is free, but clinics are busy and there is an incentive for shorter consultations. The use of eHealth software to support the integration of PRO could reduce the number of doctor visits, as reported by Kjaer et al [10] without compromising patient care. In addition, for web-based platforms used in health settings and within most practice, software has the ability to import data from other programs via interoperable formats such as XML. We achieved this using REDCap as a proxy for practice software. This served as a good test case for import and export functionalities, and the software was designed so that developing and plugging export adaptors for specific and proprietary clinical software packages is possible.

Our software was tested in a research environment, and although the esthetics were carefully considered to make it pleasurable to use, we do not know if patients will experience software *fatigue* over time, as is evidenced by the large number of mobile apps that have been designed [32]. However, we believe that a design that facilitates an experience that feels personal, in contrast with a well-constructed but impersonal form, will influence a patient's perception of the quality of care they receive in the absence of direct contact; this will be advantageous for people living in rural and remote settings. Our participants were largely computer literate, which may affect the generalizability of the product suitability for other demographic groups. In addition, some might express concern that variation in the mode of delivery might affect responses; however, research has found that when sampling protocols are followed, data equivalence can be achieved, although this research came with caveats [33]. The doctors also showed interest in using the software; however, further work will be necessary, as well as education and training on the utility of PRO data per se for patient care, as has been flagged by others [27,34]. A systematic review of mHealth adoption by Gagnon et al [35] found that utility and ease of use were two of the most important factors influencing uptake by health care professionals, and more recent reviews [30] reported organizational factors that must be considered. However, these factors vary according to the context.

Conclusions

We recognize that to be an effective tool and to provide clear value, the therapeutic benefit of using this application needs to be evident to doctors and patients [16]. This could be considered both a strength and a weakness of our design; however, the concept of the therapeutic relationship between clinician and patient underpins our design. For example, interaction with the doctor during the consultation could motivate the patient to input PRO data at home in preparation for the next visit, and equally, the clinical *return* achieved will motivate the doctor to use the software repeatedly. We also see the potential for the integration of a range of PRO instruments in a single platform, not only confined to the management of HIV infection, including, for example, side-effect questionnaires. These could help quantify the qualitative impacts of pharmaceutical interventions in tandem with medication adherence between visits. To establish the utility of this application, we envisage that a randomized controlled trial design, informed by an implementation strategy and coupled with a qualitative analysis evaluating the software over an extended period, will be necessary.

Acknowledgments

We are grateful to the participants of this study for their time and effort contributing to the research and the final product. We also thank our colleagues from the following institutions: In Western Australia: The Institute for Immunology and Infectious Diseases at Murdoch University: Mina John, Ian Bradley, Paul Egan, Ramesh Ram, Maeve Berry; and GP on Beaufort in Mt Lawley (Belinda Wozencroft, Fergus McCabe, Ric Chaney. In France: URC ECO, PRO Unit AP-HP: AP-HP: Christophe Lalanne, Jim Jouannigot, Lihui XU, Olivier Chassany, Pascal Bessonneau, Pierre Sarrot; Hôpital de Kremlin Bicetre: Cécile Goujard, Yann Quertainmont, Françoise Chouraqui, Centre Hospitalier Saint Nazaire: Nathalie Pouhaut, Carole Grand-Courault; L'Institut Mutualiste Montsouris: Loic Bodard, Hospital Lariboisiere: Agathe Rami. This work was supported by a GILEAD Sciences Europe Limited goSHAPE grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Moodboard.
[DOCX File , 1159 KB - jmir_v23i7e27861_app1.docx]

References

XSL•FO
RenderX

1. Herrmann S, McKinnon E, Hyland NB, Lalanne C, Mallal S, Nolan D, et al. HIV-related stigma and physical symptoms have a persistent influence on health-related quality of life in Australians with HIV infection. Health Qual Life Outcomes 2013 Apr 08;11(1):56 [FREE Full text] [doi: 10.1186/1477-7525-11-56] [Medline: 23566318]
2. Chan EK, Edwards TC, Haywood K, Mikles SP, Newton L. Implementing patient-reported outcome measures in clinical practice: a companion guide to the ISOQOL user's guide. Qual Life Res 2019 Mar 17;28(3):621-627 [FREE Full text] [doi: 10.1007/s11136-018-2048-4] [Medline: 30448911]
3. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res 2012 Oct;21(8):1305-1314. [doi: 10.1007/s11136-011-0054-x] [Medline: 22048932]
4. Austin E, LeRouge C, Hartzler AL, Segal C, Lavallee DC. Capturing the patient voice: implementing patient-reported outcomes across the health system. Qual Life Res 2020 Feb 12;29(2):347-355 [FREE Full text] [doi: 10.1007/s11136-019-02320-8] [Medline: 31606809]
5. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
6. Holzner B, Giesinger JM, Pinggera J, Zugal S, Schöpf F, Oberguggenberger AS, et al. The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring. BMC Med Inform Decis Mak 2012 Nov 09;12(1):126 [FREE Full text] [doi: 10.1186/1472-6947-12-126] [Medline: 23140270]
7. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
8. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. In: World Health Organisation 2019. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2021-07-27]
9. Wintner LM, Giesinger JM, Zabernigg A, Rumpold G, Sztankay M, Oberguggenberger AS, et al. Evaluation of electronic patient-reported outcome assessment with cancer patients in the hospital and at home. BMC Med Inform Decis Mak 2015 Dec 23;15(1):110 [FREE Full text] [doi: 10.1186/s12911-015-0230-y] [Medline: 26699708]
10. Kjær AS, Rasmussen TA, Hjollund NH, Rodkjaer LO, Storgaard M. Patient-reported outcomes in daily clinical practise in HIV outpatient care. Int J Infect Dis 2018 Apr;69:108-114 [FREE Full text] [doi: 10.1016/j.ijid.2018.02.015] [Medline: 29476900]
11. Schougaard LM, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de TA, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534 [FREE Full text] [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
12. Scotté F, Minvielle E, Mir O, André F, Barlesi F, Soria J. A patient reported outcome platform, a useful tool to improve monitoring and effective management of Covid-19-positive patients with cancer. Eur J Cancer 2020 Jun;132:1-4 [FREE Full text] [doi: 10.1016/j.ejca.2020.03.020] [Medline: 32294611]
13. Fredericksen R, Crane PK, Tufano J, Ralston J, Schmidt S, Brown T, et al. Integrating a web-based, patient-administered assessment into primary care for HIV-infected adults. J AIDS HIV Res 2012 Feb;4(2):47-55 [FREE Full text] [doi: 10.5897/jahr11.046] [Medline: 26561537]
14. Alyahya S, Ivins WK, Gray WA. Raising the awareness of development progress in distributed agile projects. J Software 2013 Dec 01;8(12):3066-3081. [doi: 10.4304/jsw.8.12.3066-3081]
15. Shore J, Warden S. The Art of Agile Development: Pragmatic Guide to Agile Software Development. Sebastopol, California: O'Reilly Media Inc; 2007.
16. Steele GC, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
17. Lovibond S, Lovibond P. Manual for the Depression Anxiety Stress Scales. 2nd Ed. Sydney: Psychology Foundation; 1995.
18. Duracinsky M, Herrmann S, Berzins B, Armstrong A, Kohli R, Le Coeur S, et al. The development of PROQOL-HIV: an international instrument to assess the health-related quality of life of persons living with HIV/AIDS. J Acquir Immune Defic Syndr 2012 Apr 15;59(5):498-505. [doi: 10.1097/QAI.0b013e318245cafe] [Medline: 22205438]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
22. The R Project for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/ [accessed 2021-07-09]

XSL•FO
RenderX

23. Shaffer E, Lahiri A. The Institutionalization of UX: A Step-by-Step Guide to a User Experience Practice. 2nd Ed. Upper Saddle River, New Jersey: Addison-Wesley; 2014.
24. Duracinsky M, Jouannigot J, Grace A, Power B, Desai R, Lalanne C, et al. Acceptability of electronic patient reported outcomes in clinical care of HIV. Value in Health 2016 Nov;19(7):A702. [doi: 10.1016/j.jval.2016.09.2040]
25. Duracinsky M, Lalanne C, Goujard C, Herrmann S, Cheung-Lung C, Brosseau J, et al. Electronic versus paper-based assessment of health-related quality of life specific to HIV disease: reliability study of the PROQOL-HIV questionnaire. J Med Internet Res 2014 Apr 25;16(4):e115 [FREE Full text] [doi: 10.2196/jmir.3330] [Medline: 24769643]
26. National Update on HIV, Viral Hepatitis and Sexually Transmissible Infections in Australia: 2009–2018. Sydney, NSW: The Kirby Institute for Infection and Immunity in Society; 2020:1-36.
27. Fredericksen RJ, Fitzsimmons E, Gibbons LE, Loo S, Dougherty S, Avendano-Soto S, et al. How do treatment priorities differ between patients in HIV care and their providers? A mixed-methods study. AIDS Behav 2020 Apr 12;24(4):1170-1180 [FREE Full text] [doi: 10.1007/s10461-019-02746-8] [Medline: 31832854]
28. Nielsen J. 10 usability heuristics for user interface design. Nielsen Norman Group. 1994. URL: https://www.nngroup.com/articles/ten-usability-heuristics [accessed 2021-05-18]
29. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
30. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. J Patient Rep Outcomes 2018 Dec 3;2(1):46 [FREE Full text] [doi: 10.1186/s41687-018-0072-3] [Medline: 30363333]
31. Roberts NA, Alexander K, Wyld D, Janda M. Factors in randomized controlled trials reported to impact the implementation of patient-reported outcome measures into routine care: protocol for a systematic review. JMIR Res Protoc 2019 Nov 26;8(11):e14579 [FREE Full text] [doi: 10.2196/14579] [Medline: 31769763]
32. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014 Feb;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
33. Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison C, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042 [FREE Full text] [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
34. Fredericksen RJ, Tufano J, Ralston J, McReynolds J, Stewart M, Lober WB, et al. Provider perceptions of the value of same-day, electronic patient-reported measures for use in clinical HIV care. AIDS Care 2016 Nov 29;28(11):1428-1433 [FREE Full text] [doi: 10.1080/09540121.2016.1189501] [Medline: 27237187]
35. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

Abbreviations

DASS: Depression, Anxiety and Stress Scale
HRQL: health-related quality of life
IT: information technology
mHealth: mobile health
PRO: patient-reported outcome
PROQOL-HIV: Patient Reported Outcomes Quality of Life-HIV
REDCap: Research Electronic Data Capture
WA: Western Australia

Edited by R Kukafka; submitted 22.02.21; peer-reviewed by S Mikles, I Zimianiti; comments to author 24.03.21; revised version received 19.05.21; accepted 14.06.21; published 30.07.21.

Please cite as:
Herrmann S, Power B, Rashidi A, Cypher M, Mastaglia F, Grace A, McKinnon E, Sarrot P, Michau C, Skinner M, Desai R, Duracinsky M
Supporting Patient-Clinician Interaction in Chronic HIV Care: Design and Development of a Patient-Reported Outcomes Software Application
J Med Internet Res 2021;23(7):e27861
URL: https://www.jmir.org/2021/7/e27861
doi:10.2196/27861
PMID:34328442

XSL•FO
RenderX

©Susan Herrmann, Brad Power, Amineh Rashidi, Mark Cypher, Frank Mastaglia, Amy Grace, Elizabeth McKinnon, Pierre Sarrot, Christophe Michau, Matthew Skinner, Renae Desai, Martin Duracinsky. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Approaches for Implementing App-Based Digital Treatments for Drug Use Disorders Into Primary Care: A Qualitative, User-Centered Design Study of Patient Perspectives

Joseph E Glass[1], MSW, PhD; Theresa E Matson[1], MPH; Catherine Lim[1], MDes; Andrea L Hartzler[2], PhD; Kilian Kimbel[1], BA; Amy K Lee[1], MPH; Tara Beatty[1], MA; Rebecca Parrish[3], MSW; Ryan M Caldeiro[3], MD; Angela Garza McWethy[3], MSW; Geoffrey M Curran[4], PhD; Katharine A Bradley[1], MPH, MD

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

[2]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[3]Kaiser Permanente Washington Mental Health & Wellness Services, Renton, WA, United States

[4]University of Arkansas for Medical Sciences and Central Arkansas Veterans Healthcare System, Little Rock, AR, United States

Corresponding Author:
Joseph E Glass, MSW, PhD
Kaiser Permanente Washington Health Research Institute
1730 Minor Ave
Suite 1600
Seattle, WA, 98101-1466
United States
Phone: 1 734 678 3709
Email: Joseph.E.Glass@kp.org

Abstract

Background: Digital interventions, such as websites and smartphone apps, can be effective in treating drug use disorders (DUDs). However, their implementation in primary care is hindered, in part, by a lack of knowledge on how patients might like these treatments delivered to them.

Objective: This study aims to increase the understanding of how patients with DUDs prefer to receive app-based treatments to inform the implementation of these treatments in primary care.

Methods: The methods of user-centered design were combined with qualitative research methods to inform the design of workflows for offering app-based treatments in primary care. Adult patients (n=14) with past-year cannabis, stimulant, or opioid use disorder from 5 primary care clinics of Kaiser Permanente Washington in the Seattle area participated in this study. Semistructured interviews were recorded, transcribed, and analyzed using qualitative template analysis. The coding scheme included deductive codes based on interview topics, which primarily focused on workflow design. Inductive codes emerged from the data.

Results: Participants wanted to learn about apps during visits where drug use was discussed and felt that app-related conversations should be incorporated into the existing care whenever possible, as opposed to creating new health care visits to facilitate the use of the app. Nearly all participants preferred receiving clinician support for using apps over using them without support. They desired a trusting, supportive relationship with a clinician who could guide them as they used the app. Participants wanted follow-up support via phone calls or secure messaging because these modes of communication were perceived as a convenient and low burden (eg, no copays or appointment travel).

Conclusions: A user-centered implementation of treatment apps for DUDs in primary care will require health systems to design workflows that account for patients' needs for structure, support in and outside of visits, and desire for convenience.

(J Med Internet Res 2021;23(7):e25866) doi:10.2196/25866

KEYWORDS

user-centered design; qualitative; drug use disorders; stimulants; cannabis; opioids; primary care; mHealth; mobile phone

XSL•FO
RenderX

Introduction

Background

Drug use disorders (DUDs) are prevalent and deadly worldwide [1,2]. Addiction epidemics are worsening—in 2018, 184 people a day died from drug overdose in the US, and the number of fatal-and nonfatal overdoses increased in 2020 [3,4]. However, most people with DUDs do not receive treatment [5,6]. Experts recommend that treatments for DUD should be implemented in primary care to reduce this treatment gap [7-11].

Digital treatments, such as smartphone or tablet app or websites, have been touted as a means to reach more people with effective DUD treatments [12,13]. One way to classify digital treatments is to place them on a spectrum ranging from apps that are used as self-care without the help of a clinician to apps that are fully incorporated into patients' health care and guided by a clinician [14]. Several clinician-guided treatments for alcohol or drugs are supported by the evidence of efficacy or effectiveness [15-17]. For instance, two platforms—the Therapeutic Educational System and CBT4CBT—deliver cognitive behavioral treatment through web-based modules [15,18]. Both were initially designed by researchers to augment standard care for substance use disorder and have since been marketed to health care systems in the form of apps or websites with patient-facing and clinician-facing features [19,20].

Despite the promise of digital treatments, there is a lack of knowledge about how to optimally integrate them into routine primary care. In trials of digital treatments, including those for DUD and other health conditions, patients often fail to engage with the software, leading to null results [21,22] and failed real-world clinical trials [23]. Without adequate support to enhance motivation to engage with apps, patients prematurely stop using apps or use them rarely, thereby decreasing efficacy [21,22,24-26]. This poses design and implementation problems in primary care. Evidence suggests that patients can benefit from digital treatments if they receive extensive coaching and help to use them. In contrast, the time of patients and clinical teams is limited and filled with competing demands [27-29]. Therefore, the integration of apps into health care will need to balance patients' desires and needs with these constraints.

Objectives

To incorporate patient voices into the design of approaches for offering apps for DUDs in primary care, this study combined user-centered design methods [30] with qualitative research methods. Drawing from the medical informatics literature, work systems models are often used to guide the design of clinical workflows [31-33]. Central to the concept of work systems' models are actions, people, and tools that can assist patients and health care teams in embracing health care technologies. We applied these concepts to understand patient preferences on how to introduce DUD treatment apps to patients, assist them with the app setup, and offer appropriate follow-up over time. Consistent with the principles of user-centered design [30], this study served as the first within a series of implementation science studies that will iteratively design and test the effectiveness of approaches for implementing digital treatments for DUDs in real-world health care settings.

Methods

Study Setting

Study participants were recruited from 5 primary care clinics of Kaiser Permanente Washington, a regional integrated health care system, in Seattle area. All clinics employed licensed independent clinical social workers with some training in DUD interventions, and several clinics had primary care providers (PCPs) who prescribed buprenorphine. Consultative addiction psychiatry was available in the health system. Specialist addiction treatment programs were available through an external care network that was contracted by the health system [34].

Eligibility Criteria

The main eligibility criteria, assessed by phone screening, were smartphone use to ensure a basic familiarity with these devices [35] along with the presence of a past-year cannabis, stimulant, or opioid use disorder based on the Mini International Neuropsychiatric Interview for the *Diagnostic and Statistical Manual of Mental Disorders*–Fifth Edition [36,37]. During our accrual period, we revised the eligibility criteria to also include patients without a past-year opioid use disorder based on *Diagnostic and Statistical Manual of Mental Disorders*–Fifth Edition if they had been prescribed buprenorphine formulations used to treat opioid use disorder [38].

Participant Identification and Recruitment

Between July 2018 and December 2018, we queried electronic health records to identify patients aged 18 years or more with a recent primary care visit, who had screened positive for cannabis or other drugs [39-42] or had a documented DUD diagnosis. The study staff mailed invitations with a preincentive of US $2 to 101 potentially eligible patients [43,44]. We phoned 77 of these patients for eligibility screening while providing a US $10 incentive for those who completed phone screening, regardless of eligibility. Multimedia Appendix 1 presents a participant recruitment flow diagram noting the reasons for exclusion. We used purposeful sampling [45] to promote sample diversity by race, gender, age, substance type (opioid, cannabis, stimulant), and prior DUD treatment or mutual support program attendance [46].

Sample

The study's initial recruitment goals were based on pragmatic and empirical criteria. We sought to capture a sufficient diversity of patient perspectives to inform the initial implementation of digital treatments in a primary care practice that could then be iteratively adjusted during implementation within a health care system. We targeted a minimum of 12 participants [47] and stopped recruitment at 14 while monitoring saturation during the analysis [48]. Among the 14 participants, 6 had opioid use disorder or were taking buprenorphine, 9 had cannabis use disorder, and 4 had stimulant use disorder (participant details in Multimedia Appendix 1).

Semistructured Interviews

The interview guide was designed to elicit preferences regarding workflow design among candidate approaches for offering and supporting the delivery of app-based DUD treatments. The

XSL•FO
RenderX

interviews utilized user-centered design tools, including personas and storyboards. Personas depicted a visual *user profile* to show participants an example of a person who might be a typical user of an app [49]. We developed three personas—one each for stimulants, opioids, and cannabis—that depicted a hypothetical patient with several DUD symptoms who was interested in learning about options for help (Multimedia Appendix 1). We presented one persona per participant. We also used storyboards [50], which used illustrated panels to show core features of candidate workflow designs.

The storyboards showed alternative scenarios to support patients' app use. Scenarios fit into three workflow phases: (1) introducing the app to patients and helping them learn about it (*Introduction*); (2) setting up patients interested in using the app with the treatment (*Setup*); and (3) following up with patients who agree to use the app to promote engagement and execute a care plan that includes the app (*Follow-up*). There were 3 to 4 candidate scenarios as options for each workflow phase (Figure 1; complete storyboards are presented in Multimedia Appendix 1, which includes detailed information about the storyboard design process and rationale behind each scenario). Scenarios varied in whether interactions were done virtually or in person, how much help the patient received, which team members were involved, and other aspects of communication and workflow. Interviewers read aloud all scenarios within a phase, asked participants to rank scenarios according to their preferences, and probed for additional information.

Figure 1. Scenarios depicting potential introduction (learning about the app), setup (getting started with the app), and follow-up (getting follow-up while using the app) phases for a drug use disorder treatment app. The boxes denote the overall workflow most preferred by participants. PCP: primary care provider.



Study interview materials were informed by domains of work systems models (eg, workflow actions or tasks, individuals or people, and tools that facilitate work) [33], implementation science frameworks (eg, the extent of support and how to guide patients) [14], and prior digital treatment studies (eg, privacy issues and concerns regarding the use of electronic self-reported data) [22,51-53]. The guide also drew from theory and literature on patient treatment engagement and health services design in primary care [21,22,52-63]. Clinical leaders and clinicians from Kaiser Permanente and a safety net health clinic provided input to the materials. Although not analyzed in this study, the interviews asked about background experiences (eg, app use; Multimedia Appendix 1 contains an interview guide summary). Interviews were conducted in person and digitally recorded and professionally transcribed. Interviews lasted for an average of 87 minutes, including informed consent.

The Kaiser Permanente Washington Institutional Review Board granted ethical approvals. Written informed consent was obtained from all the participants enrolled in this study. A waiver of written informed consent was obtained to conduct telephone

XSL•FO
RenderX

screenings. A waiver of informed consent was obtained for the identification of potential participants using electronic health records.

Data Analysis

For qualitative analysis, we imported transcripts into Dedoose v7.0.23 (SocioCultural Research Consultants) [64]. We applied template analyses to qualitatively code excerpts of text [65], which used a combination of inductive and deductive approaches. A priori codes were based on the interview guide topics. Codebook definitions were informed posthoc by the Workflow Elements Model [31,32] constructs to help in classifying participant preferences regarding the physical or virtual *actions* that need to be performed, the *people* (eg, clinicians) who need to perform the actions, and the physical and virtual *tools* used to promote app delivery and use. For the first 5 transcripts, 2 study team members coded them, a third study team member reconciled them, and the team discussed achieving consensus in the form of a codebook. To further increase rigor, we kept memos and an audit trail of code revisions [66,67]. The remaining transcripts were split among team members for coding, and the principal investigator (PI; JEG) reconciled them. The lead analyst (TEM) examined the fully coded data for patterns within the coding hierarchy and preliminary themes that occurred across codes, regularly meeting with the PI to review results. In several meetings, the PI and lead analyst used affinity mapping, a user-centered design activity that helps in clustering similar ideas or concepts together [68] to further refine the themes. The themes were discussed with other team members before they were finalized.

Availability of Data and Materials

The interview materials are provided in Multimedia Appendix 1. Additional interview materials are available from the first author upon reasonable request.

Results

Preferences Regarding Workflow Actions

This section presents patient preferences regarding the actions performed while facilitating a DUD treatment app, by each workflow stage.

Introduction Stage: Learning About the Treatment App

Nearly all participants preferred to learn about app-based treatment options during an in-person primary care visit where drug use was already being discussed. Participants generally did not want to learn about an app during an unexpected *cold* call with a clinician they had not met. Cold calls in relation to drug use would feel intrusive, potentially awkward, and embarrassing. Participants who were open to receiving a phone call to learn about an app said they would at least need to know about the phone call in advance. Only 2 participants were interested in a self-guided treatment where they would learn about an app via a pamphlet without any interaction with a health care professional, and only 1 participant said that they would respond to a waiting room advertisement. A few noted that the convenience of accessing the app without help from a clinician was appealing, but others said they would not be motivated to use an app without support. As one participant explained:

> *I just don't see that I would pick [the flyer] up ... It would be more effective if, after meeting with my primary care person about my addiction issues, they got me in direct contact with [a mental health clinician].* [P3, who was receiving medication treatment for opioid use disorder]

Setup Stage: Getting Started With the App

Overall, participants thought that it was critical to avoid getting "*stuck*" while setting up and starting to use an app. Having a seamless experience would be critical for maintaining motivation. One participant said the following:

> *Sit next to me and walk me through it. Hold my hand.* [P13, participant with cannabis use disorder]

Most participants preferred in-person assistance to help them get started with the app once they decided to use it. However, several others wanted to receive this help over the phone to reduce the amount of time spent at their visit or to avoid returning for a visit. For these participants, the phone setup was "the best of both worlds." One participant said as follows:

> *number one is like, we agree that [persona name] will schedule an in-person appointment. In my case I'm very reluctant to do stuff like that, so okay, extra work for me–awesome, thanks. This one [solo setup scenario] is like all on my own, no support. So this one [phone setup scenario] I feel like is the best of both worlds.* [P3, participant with cannabis use disorder]

Still, some thought that scheduling phone appointments would be a *chore*. Few preferred to set up the app without help; these participants noted that setting up the app on their own would save time.

Follow-up Stage: Getting Support Over Time for Engagement and Care With an App

Half of the participants preferred follow-up to occur over the phone. This would provide personalized communication and a relationship that would help them stay accountable without the inconvenience and monetary costs of an in-person visit. Even those who favored an initial visit in person generally preferred a phone follow-up. Few preferred a follow-up strategy that only involved secure messages, but they liked the idea of receiving messages in addition to phone calls. One participant noted:

> *Secure messages and phone for me ... Because that way you have somebody on the phone to tell you if you've interpreted the message right ... I've had that happen many times. A doctor will say something and I don't necessarily interpret it correctly or the way that it's meant.* [P14, participant with opioid and stimulant use disorder]

Most participants thought that reading and responding to a message at a later time would be convenient, and some were more receptive to more frequent contacts if done over messages. However, others noted that they do not check messages or

XSL•FO
RenderX

thought they would be "too easy to ignore." One participant added the following:

> *If you're having that reluctance, that's so much easier to just be like well, no, I'm not going to do that.* [P10, participant with cannabis use disorder]

Preferences Regarding the People Who Could Deliver Digital Treatments

Preferences regarding the people who could facilitate a DUD treatment app were often applicable across multiple stages of the workflow (Multimedia Appendix 1). Thus, the preferences are presented here for the overall workflow.

Most participants preferred a clinician with mental health expertise to introduce and guide them in app use over time. Several participants said that mental health clinicians might have more knowledge about treatment options than a PCP, more experience treating drug use, or more time to describe features of an app and walk them through the setup process. However, some preferred to only talk to PCPs and said that mental health clinicians would make it feel like their "issue was serious." Other benefits of working with a PCP ranged from wanting to work with someone with whom they had an established rapport, wanting fewer clinicians involved in their care, and wanting to work with someone who can offer medical advice. Many acknowledged logistical constraints of working with a PCP (eg, limited time and challenge of scheduling appointments). One participant said the following:

> *I don't think the primary care [provider] needs to be concerned with that. They need to have some knowledge of it, but I think the major, the main focus of it would be with the social worker, plus the social worker would be able to follow up with the patient as far as once they get started using the app. It would be easier follow up for the social worker than the primary care [provider].* [P5, participant with stimulant use disorder]

A few participants said it would be helpful to hear from someone with lived experience who had "been through it before" to help them decide whether to use the app over other treatment options. They did not feel that it was important to have such an individual remain involved after they had started using the app.

Finally, participants described the value of having access to technical support for help with downloading, setup, and use of the app. Participants pointed out that they would not want technical issues to consume valuable visit time. Participants clearly differentiated technical or setup assistance from treatment-related assistance and drew boundaries around the type of help they would want from each person.

Preferences Regarding the Tools That Could Help Facilitate the Use of the Digital Treatment

Participants described a range of ideas about virtual or physical tools that could facilitate the delivery of an app (Textbox 1). For instance, pamphlets, user ratings and reviews, and trial versions of the app were suggested as tools for the introduction stage. Written instructions and video tutorials could help with the setup stage. Several tools were suggested for the follow-up stage to facilitate communication (eg, a *"Get Support"* button to help contact the care team).

> *Having a way to check in with your doctor about it would be useful ... In [the phone setup scenario] if there could be a way to demonstrate it like in [the in-person scenario], with a screen sharing thing – I think that would be super useful, or maybe like a video, like an instructional video showing how to navigate it, that would be useful, since you wouldn't be able to do it in person.* [P12, participant with cannabis and stimulant use disorder]

XSL•FO
RenderX

Textbox 1. Tools that could help facilitate the use of a drug use disorder treatment app in primary care, as suggested by patients. Tools are presented by the workflow stage.

Introduction Stage: Learning About the Treatment App

- Pamphlet explaining how the treatment app will benefit the patient
- User ratings and reviews that could lend credibility to the treatment app
- Trial version or demonstration of the treatment app allowing patients to test it out before committing
- Advertisements or information about treatment app on a health plan or agency website

Setup Stage: Getting Started With the App

- Written instructions for getting started with the treatment app or a user guide
- Video tutorial for getting started with the treatment app
- Smartphone requirements for downloading and running the treatment app (memory storage, operating system version, and needing to know their app store password)
- A webpage or button in the app that provides answers to frequently asked questions

Follow-up Stage: Getting Support Over Time for Engagement and to Execute a Care Plan While Using the App

- Telephone caller identification (*caller ID*) so that the patient knows if their clinician is calling (patients may not answer their phone if the incoming call looks like a generic or toll-free number)
- Contact or *Get Support* button
- Technical assistance (for app and device-related questions)
- Clinician contact information or direct messaging feature (for treatment questions)
- Screen sharing functionality so the patient's smartphone screen can be viewed by a support professional
- Reminders and notifications, sent in accordance with the level of engagement with the treatment app (ie, more frequent reminders or notifications if patient is not using the app)

Cross-Cutting Themes

Four cross-cutting themes emerged across codes in the codebook. Themes are described below; Textbox 2 provides representative quotes for each theme (Multimedia Appendix 1).

XSL•FO
RenderX

Textbox 2. Themes derived from analyses across the codebook: cross-cutting recommendations for designing a patient-centered approach for offering treatment apps for drug use disorders in primary care

Established Relationships and Trust Would Facilitate a Better Patient Experience

- "It's a person who's already talking to her about her drug addiction, like supplying Suboxone or whatever, that is saying hey, this might help. And I feel like people would be a little bit more receptive if it's someone that they already trust with their treatment." (P1, participant with cannabis and opioid use disorder. Had prior drug use disorder [DUD] treatment.)
- "I mean the main thing is support...I would say the best people that will help you is if they understand what you're going through, especially during that period, and they need to understand what you went through. It doesn't necessarily have to be that way, but that's how I was, and primary care provider helped me in the professional way, and then the emotional issues of that time...But I will say if someone is truly trying to help this person making them know that there's still someone that cares about them, whatever, in any way it's a good thing. It is. Some people have no one, and even small interactions can make a difference in positive ways." (P2, participant taking buprenorphine whose opioid use disorder is in remission. Had prior DUD treatment.)
- "I know for me personally that if I'm talking with somebody–I mean, if you can send me something, I can more likely put it on the back burner. Whereas if I have an actual conversation with them – I don't know, it's just more personal. I think it's more of a personal check in, that there's a real human being right there that's interested in my care." (P3, participant taking buprenorphine whose opioid use disorder is in remission. Had prior DUD treatment.)

Patients Were Open to Team-Based Approach When Getting Support From Their Primary Care Team in the Use of Apps

- "I mean if there were actually social workers doing this, then I suppose that would be who I would be contacting, and I wouldn't really have any issue with that because that's just sort of how it would be. It would be another person I'm meeting to satisfy a different medical need that I have. I don't have a lot of hang-ups about meeting different providers for different issues. I've done a lot of that in the past." (P8, participant with cannabis and stimulant use disorder. Had prior DUD treatment.)
- "I see the benefit of having somebody else who's maybe more focused on either the treatment or the app itself...I don't think the primary care [provider] needs to be concerned with that. They need to have some knowledge of it, but I think the major, the main focus of it would be with the social worker, plus the social worker would be able to follow up with the patient as far as once they get started using the app. It would be easier follow up for the social worker than the primary care." (P5, participant with stimulant use disorder. Had prior DUD treatment.)

Patients Felt a Tension Between Effectiveness and Convenience in Aspects of Workflow Design

- "Frankly I don't have a lot of time, so if I could do it over the phone I would...But I think this one [an in-person visit] would be more effective for a lot of people." (P2, participant taking buprenorphine whose opioid use disorder is in remission. Had prior DUD treatment.)
- "Probably I have a slight preference for an in-person visit, just on the basis that I definitely have an easier time talking with someone in-person than over the phone. But yeah, in-person would work a little better in that sense, but on the phone is also very convenient." (P10, participant with cannabis use disorder. No prior DUD treatment).

The Workflow Needs to Meet Patients Where They Are At

- "The first [phone call] I'd even say like within three days. Because if I go to the doctor and you gave me a screening and realize I'm an addict, you give me this thing, I actively want to make a difference, and the next day ...hang out with some friends and do coke, I'm not going to remember that. So two or three days later I would be totally–I wouldn't be annoyed by that...And then like every week after that, just as a – hey, I'm here. I'm that app, remember? But the likelihood of me actually going home and doing it right away, probably very little. I would either have to get worse in whatever I was doing, or maybe the day after, on that terrible hangover, you're like God, I need help – that's when I'd probably look at it, or like look for it and try and find it." (P4, participant with cannabis and stimulant use disorder. No prior DUD treatment.)
- "if it seems like I'm on track and using the app more often, then I wouldn't need as many reminders, but if I'm off track, then it would be more helpful for someone is keeping up with me more constantly to hold me to it." (P10, participant with cannabis use disorder. No prior DUD treatment.)

Established Relationships and Trust

Participants told us that they placed value in having a connection with the health care professional who worked with them on using a digital treatment. It would be ideal if this would be someone who had already established a relationship with them, regardless of their clinical role. Even if there was no pre-existing relationship, it was critical for this person to be compassionate and caring because conversations about substance use can be stigmatizing or embarrassing.

Openness to a Team-Based Approach

Although participants had a desire to work with someone with whom they had an existing relationship, the trust between a PCP and a patient appeared to extend to the broader primary care team. That is, a PCP would not necessarily be the one *holding [their] hand* throughout using an app. Some noted that PCPs are busy and often delegate care to other team members, including DUD care. However, there were some important bounds around this division of responsibility. For instance, having a "third party" from outside the health system support patients was seen as a bad idea.

An important exception to this theme is that a participant firmly wanted to work with one clinician. Needing to talk to an additional person about an app could be an "extra step" and could "add like the risk that they won't do it at all."

XSL•FO
RenderX

Tension Between Effectiveness and Convenience

Participants described the pros and cons of the intensity of hypothetical workflow interactions. Regarding the general workflow design, they noted that convenience and effectiveness were in opposition. In-person visits were more "personable" and would hold them more "accountable" to using an app; however, an in-person visit would require more time, travel, and potentially copays. Secure messages, or receiving reminders through the app, were considered convenient because participants could read and respond whenever they wanted; however, they acknowledged that these messages might not be as powerful as a phone call or face-to-face visit in keeping them engaged.

The Workflow Needs to Meet the Patient Where They Are At

Participants suggested that workflows should be tailored to individuals depending on their level of motivation, how often they use substances, and how long or how successfully they have been managing their substance use. For instance, this desire for tailoring reflected that someone with less motivation may need more hands-on help, someone who is using substances daily might want to be contacted every day, and someone who is just beginning to address their use and/or who is actively struggling may need more frequent follow-ups.

Synthesis

Upon synthesis of the qualitative data, and considering the number of participants who preferred each storyboard scenario (Multimedia Appendix 1), we suggest a general approach for facilitating the use of DUD treatment apps, which is described using the story of a fictitious patient (Table 1). Briefly, once the patient expresses interest in using an app for DUD, the PCP confirms their willingness to talk with a clinician with mental health expertise that will help them get started and keep them accountable while initiating treatment. The clinician would teach the patient about the app, help them plan for using the app, agree on a structure for follow-up, and then provide regular follow-up by telephone and messaging (bold boxes, Figure 1). The clinician would further build rapport and ask the patient their preferred cadence, method of follow-up, and modality of contact (eg, in person vs phone). Importantly, the specific approach would remain flexible and tailored to the patient.

Table 1. A general approach for supporting patients in using a drug use disorder treatment app in primary care, based on participant preferences.

Workflow stage	Hypothetical experience for a fictitious patient *Cory*	Why participants liked this experience	Other experiences preferred by participants
Introduction stage: learning about the treatment app	Cory completes an annual health screen that asks about alcohol and drug use. Cory's PCP[a] expresses concern that her regular substance use could affect her health. Cory is interested in learning about options that could help her change. She agrees to talk to a mental health clinician on the primary care team. Privately, Cory and the mental health clinician discuss Cory's goals for change, and review a few different options, including a treatment app for drug use.	Participants wanted to discuss substance use with a provider they already knew—but they also recognized that their PCP might not have the time or right expertise. Being seamlessly connected to a mental health clinician would feel *"more personal"* and provide support beyond what their PCP could offer.	Participants who wanted to talk to only one person—usually their PCP—said they would try to set up the treatment app on their own after their PCP ordered it.
Setup stage: getting started with the app	A mental health clinician describes features and content of the DUD[b] treatment app that might be helpful to Cory. They give Cory instructions for how to get started, and they agree to check in after a couple weeks.	Most participants wanted to learn from someone on their care team how an app would benefit them and how to use it. Chances of using the app would be higher if these were discussed when motivation was high.	Several participants felt comfortable getting started with a treatment app on their own. Some said that technical support would be necessary.
Follow-up stage: getting support over time for engagement and executing a care plan while using the app	Cory gives the app a try. She eventually stops using the app after a couple weeks. However, she re-engages with the app after exchanging secure messages with her mental health clinician that covers a status update, tailored recommendations for using the app, and plans for a follow-up phone call.	Participants said that phone follow-up offered more support than secure messages and placed fewer demands on their time or finances than an in-person visit. Many wanted follow-up spaced out over time to help hold them accountable to using the app.	Benefits of follow-up via secure message include the choice in when and how often to responded to messages (unlike visits or phone calls). In-person appointments would be reserved for additional support and accountability.

[a]PCP: primary care provider.

[b]DUD: drug use disorder.

Discussion

Principal Findings

This study conveyed patient perspectives on the use of apps as part of treatment for DUD in primary care. Overall, participants desired to receive support from their health care teams in using apps and voiced little interest in using them without clinician guidance. There was a consensus among participants that they preferred to work with a trusted, competent clinician who could guide them in using the app over time. They stressed the importance of follow-up and felt that in most cases, this could

XSL•FO
RenderX

generally be done through telephone with the addition of asynchronous secure messaging whenever needed.

These findings contribute novel information about patient preferences, laying the groundwork for research on the implementation of apps for DUDs in health care. Patients desire low-barrier, nonstigmatizing interventions for DUD in primary care [69,70]. Apps could potentially help address this gap in care. Although prior clinical trials have relied on research staff to "*train*" patients in using apps or facilitate their ongoing use [15,21,71-74], future studies can use these findings to inform the involvement of health care teams, instead of researchers, in various aspects of app delivery and implementation.

Overall, participants expressed that the most ideal way to offer apps for DUDs was during routine clinical interactions regarding drug use. This approach enables several aspects of care that participants desired, including the ability to build trust with a clinician, obtain medical and mental health advice, and go home with an appropriate app and clear expectations about treatment. The literature has described an alternative approach, where apps are offered as a self-guided option in the absence of clinician guidance [14,75], potentially reducing health care costs and burden. Such an approach was generally not preferred by participants in this study; some wanted access to a self-care option, but clinician-guided care was viewed as more effective and engaging. The findings are consistent with the literature that describes the need for *supportive accountability* while delivering app-based treatments, where patients establish a relationship with a helper that, among other things, sets expectations for app use and provides assistance over time [76]. We note that this study excluded participants whose drug use did not lead to a DUD; perhaps, self-care options should be further studied among patients with lower drug use severity.

Participants felt that treatment with an app should be seamless, free of technical glitches, and other barriers that could decrease their motivation or lead to "*getting stuck*." Apps and associated clinical workflows need to impose few barriers. Indeed, many people with DUD report logistical barriers to treatment [77-82]. Some participants were also concerned about copays, which contributed to their support preferences. Future studies should design app implementation approaches that address socioeconomic and other barriers that could lead to the inequitable provision of these treatments [83,84].

The literature has also highlighted potential logistical challenges from a health system perspective. For instance, PCPs are heavily burdened and busy [27-29]. Indeed, a prior implementation study of a DUD app had to modify its initial plans by not involving PCPs [51]. Fortunately, participants in this study acknowledged these difficulties and were open to working with team members other than their PCP. Some studies in primary care failed to adequately engage patients in app use over time [21,22], and/or have been halted because workflows were not adequately developed to reach and communicate with patients [23]. This study advances the literature by adding patient voices that can inform future research on app delivery and implementation in routine care.

Limitations

This study has several limitations. The sample size was small, which limits generalizability; however, we followed the standard of saturation in qualitative research. Furthermore, our use of purposeful sampling promoted a rich accounting of a diversity of patient perspectives and preferences across important demographic and substance use characteristics to inform the implementation of apps. Nonetheless, the small sample size limited our ability to identify the ideas and preferences of all possible patients. Although there was saturation in the analysis of workflow design principles and cross-cutting themes (eg, new data analyzed were repetitious of prior data and there was no indication of new emerging themes), the ideas regarding *tools* remained diverse. In this case, we opted to present a table that captured all participants' ideas. Future study iterations are driven by additional data collection. Participants were from 5 clinics of a single integrated health care system in the United States from a single geographic region. Participants may have been more accustomed to working with multiple care providers than patients who received care in health systems with little care coordination. Thus, the results might not translate to patients from other groups. Although the goal of this study was to understand patient perspectives, clinicians' voices were not assessed, which is also a limitation. This study's focus on patients was driven by knowledge that variations in patient engagement impacted the results of trials of digital treatments [21,22,24-26,73]. Future research should elicit clinician perspectives on the mechanics of delivering apps for DUDs in primary care.

Conclusions

The perspectives of primary care patients with cannabis, opioid, and stimulant use disorders suggest that a user-centered implementation of DUD treatment apps in primary care will require health systems to guide and support patients. Research is needed to evaluate clinician perspectives on workflows for delivering apps and to test the feasibility of the design considerations suggested by participants in this study. One such study is our ongoing implementation research trial funded by the National Institute on Drug Abuse (award number R01DA047954), which uses a randomized factorial design to evaluate different approaches for implementing a digital treatment for substance use disorders, while concurrently collecting qualitative data on implementation and workflow design. As more emphasis is placed on the use of digital treatments for DUDs for primary care patients, convenient methods for engaging patients and supporting them before, during, and after treatment will be paramount.

Acknowledgments

The authors are grateful to the patients of Kaiser Permanente Washington for volunteering for this study. The authors would also like to thank Dr Manu Thakral at the University of Massachusetts Boston for providing input on workflow appropriateness and

XSL•FO
RenderX

feasibility and Kristen Delaney and Kay Theis at Kaiser Permanente Washington Health Research Institute for their expert database and electronic medical records programming that made this study possible. Finally, the authors thank the faculty and fellows of the National Institutes of Health Implementation Research Institute and the Mixed Methods Research Training Programs for their feedback on this work. The primary funding for this study was obtained through a pilot award to JEG from the Kaiser Permanente Washington Research Institute Development Fund. JEG and KAB were awarded funding by the National Institute on Alcohol Abuse and Alcoholism under K01AA023859 and K24AA022128. This was also supported by the National Institute on Drug Abuse under award R01DA047954. GMC is supported by the Translational Research Institute UL1TR003107 through the National Center for Advancing Translational Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

Funding acquisition and the first manuscript draft was done by JEG. Study conceptualization and design was conducted by JEG, TEM, CL, ALH, RP, RMC, AGM, GMC, and KAB. Data acquisition was carried out by JEG, TEM, CL, KK, AKL, TB, RP, and RMC. Data analysis was done by TEM, JEG, CL, AKL, and GMC. All authors were involved in editing and manuscript approval.

Conflicts of Interest

Several authors (JEG, RP, RMC, KAB) are investigators on a trial funded by the National Institute on Drug Abuse (R01DA047954; PI: Glass), which evaluates the implementation of reSET and reSET-O that are digital therapeutics for substance use disorder marketed by Pear Therapeutics, Inc. Pear Therapeutics Inc provided digital therapeutic prescriptions at no cost to Kaiser Permanente Washington during a quality improvement pilot study that precedes the implementation trial. JEG is a coinvestigator on a Small Business Innovation Award funded awarded to Pear Therapeutics, Inc by the National Institute on Drug Abuse, which evaluates potential improvements to reSET-O (R44DA042652). This study is informing our research on the implementation of digital therapeutics, but Pear Therapeutics, Inc was not involved in this study and does not provide funding to the authors.

Multimedia Appendix 1
Interview materials and additional results for the study. Approaches for implementing app-based digital treatments for drug use disorders into primary care: A qualitative, user-centered design study.
[DOCX File , 2001 KB - jmir_v23i7e25866_app1.docx]

References

1. Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Shirude S, Unützer J, et al. Trends and patterns of geographic variation in mortality from substance use disorders and intentional injuries among US counties, 1980-2014. J Am Med Assoc 2018 Mar 13;319(10):1013-1023 [FREE Full text] [doi: 10.1001/jama.2018.0900] [Medline: 29536097]
2. US Burden of Disease Collaborators, Mokdad AH, Ballestros K, Echko M, Glenn S, Olsen HE, et al. The state of US Health, 1990-2016: burden of diseases, injuries, and risk factors among US states. J Am Med Assoc 2018 Dec 10;319(14):1444-1472. [doi: 10.1001/jama.2018.0158] [Medline: 29634829]
3. Hedegaard H, Miniño AM, Warner M. Drug overdose deaths in the United States, 1999–2018. Center for Disease Control and Prevention. Atlanta, GA: Center for Disease Control and Prevention; 2020. URL: https://www.cdc.gov/nchs/products/databriefs/db356.htm [accessed 2020-09-01]
4. The consequences of COVID-19 on the overdose epidemic: overdoses are increasing. Overdose Detection Mapping Application Program (ODMAP). 2020. URL: http://odmap.org/Content/docs/news/2020/ODMAP-Report-May-2020.pdf [accessed 2021-05-24]
5. Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, et al. Epidemiology of DSM-5 alcohol use disorder: results from the National Epidemiologic Survey on alcohol and related conditions III. JAMA Psychiatry 2015 Aug;72(8):757-766 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.0584] [Medline: 26039070]
6. Grant BF, Saha TD, Ruan WJ, Goldstein RB, Chou SP, Jung J, et al. Epidemiology of DSM-5 drug use disorder: results from the National Epidemiologic Survey on alcohol and related conditions-III. JAMA Psychiatry 2016 Jan;73(1):39-47 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.2132] [Medline: 26580136]
7. Ghitza UE, Tai B. Challenges and opportunities for integrating preventive substance-use-care services in primary care through the Affordable Care Act. J Health Care Poor Underserved 2014 Feb;25(1 Suppl):36-45 [FREE Full text] [doi: 10.1353/hpu.2014.0067] [Medline: 24583486]
8. Saitz R, Larson MJ, Labelle C, Richardson J, Samet JH. The case for chronic disease management for addiction. J Addict Med 2008 Jun;2(2):55-65 [FREE Full text] [doi: 10.1097/ADM.0b013e318166af74] [Medline: 19809579]
9. Crowley RA, Kirschner N, Health and Public Policy Committee of the American College of Physicians. The integration of care for mental health, substance abuse, and other behavioral health conditions into primary care: executive summary of an American College of Physicians position paper. Ann Intern Med 2015 Aug 18;163(4):298-299. [doi: 10.7326/M15-0510] [Medline: 26121401]

XSL•FO
RenderX

10. Institute of Medicine. Improving the Quality of Health Care for Mental and Substance-Use Conditions. Washington, D.C., United States: National Academy Press; 2006:1-204.
11. National Council for Behavioral Health. URL: https://healthinsight.org/tools-and-resources/download/123-educational-resources/1535-implementing-care-for-alcohol-other-drug-use-in-medical-settings-an-extension-of-sbirt [accessed 2021-05-24]
12. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares M, et al. Integrating addiction treatment into primary care using mobile health technology: protocol for an implementation research study. Implement Sci 2014 May 29;9(1):65 [FREE Full text] [doi: 10.1186/1748-5908-9-65] [Medline: 24884976]
13. Marsch LA, Campbell A, Campbell C, Chen C, Ertin E, Ghitza U, et al. The application of digital health to the assessment and treatment of substance use disorders: the past, current, and future role of the National Drug Abuse Treatment Clinical Trials Network. J Subst Abuse Treat 2020 Mar;112S:4-11 [FREE Full text] [doi: 10.1016/j.jsat.2020.02.005] [Medline: 32220409]
14. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
15. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. Am J Psychiatry 2014 Jun;171(6):683-690 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13081055] [Medline: 24700332]
16. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
17. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Dec 25;9:CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
18. Kiluk BD, Nich C, Buck MB, Devore KA, Frankforter TL, LaPaglia DM, et al. Randomized clinical trial of computerized and clinician-delivered CBT in comparison with standard outpatient treatment for substance use disorders: primary within-treatment and follow-up outcomes. Am J Psychiatry 2018 Sep 01;175(9):853-863 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17090978] [Medline: 29792052]
19. Pear Therapeutics, Inc., Boston. 2019 Mar 2. URL: https://peartherapeutics.com/products/reset-reset-o/ [accessed 2020-03-02]
20. Scientific evidence is at the heart of chess. Chess Health. 2021. URL: https://www.chess.health/evidence/ [accessed 2021-02-17]
21. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2015;351:h5627 [FREE Full text] [Medline: 26559241]
22. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open 2015 Nov 30;5(11):e008581. [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]
23. Thies K, Anderson D, Cramer B. Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a federally qualified health center: interview analysis of staff and patients in a failed randomized trial. JMIR Hum Factors 2017 Oct 03;4(4):e24 [FREE Full text] [doi: 10.2196/humanfactors.7709] [Medline: 28974481]
24. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
25. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
26. Dedert EA, McDuffie JR, Stein R, McNiel JM, Kosinski AS, Freiermuth CE, et al. Electronic interventions for alcohol misuse and alcohol use disorders: a systematic review. Ann Intern Med 2015 Aug 4;163(3):205-214. [doi: 10.7326/M15-0285] [Medline: 26237752]
27. Chernof BA, Sherman SE, Lanto AB, Lee ML, Yano EM, Rubenstein LV. Health habit counseling amidst competing demands: effects of patient health habits and visit characteristics. Med Care 1999 Aug;37(8):738-747. [doi: 10.1097/00005650-199908000-00004] [Medline: 10448717]
28. Jaén CR, Stange K, Nutting P. Competing demands of primary care: a model for the delivery of clinical preventive services. J Fam Pract 1994 Feb;38(2):166-171. [Medline: 8308509]
29. Rost K, Nutting P, Smith J, Coyne J, Cooper-Patrick L, Rubenstein L. The role of competing demands in the treatment provided primary care patients with major depression. Arch Fam Med 2000 Feb;9(2):150-154. [doi: 10.1001/archfami.9.2.150] [Medline: 10693732]
30. Ergonomics of Human-System Interaction - Part 210: Human-centred design for interactive systems. Industry Standards Organization. 2019. URL: https://landingpage.bsigroup.com/LandingPage/Undated?UPI=000000000030388991 [accessed 2021-05-24]

XSL•FO
RenderX

31. Mickelson RS, Unertl KM, Holden RJ. Medication management: the macrocognitive workflow of older adults with heart failure. JMIR Hum Factors 2016 Oct 12;3(2):e27 [FREE Full text] [doi: 10.2196/humanfactors.6338] [Medline: 27733331]
32. Unertl KM, Novak LL, Johnson KB, Lorenzi NM. Traversing the many paths of workflow research: developing a conceptual framework of workflow terminology through a systematic literature review. J Am Med Inform Assoc 2010;17(3):265-273 [FREE Full text] [doi: 10.1136/jamia.2010.004333] [Medline: 20442143]
33. Carayon P. The balance theory and the work system model… twenty years later. Int J Hum-Comput Int 2009 Jun 08;25(5):313-327. [doi: 10.1080/10447310902864928]
34. Glass JE, Bobb JF, Lee AK, Richards JE, Lapham GT, Ludman E, et al. Study protocol: a cluster-randomized trial implementing Sustained Patient-centered Alcohol-related Care (SPARC trial). Implement Sci 2018 Aug 06;13(1):108. [doi: 10.1186/s13012-018-0795-9] [Medline: 30081930]
35. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. JMIR Mhealth Uhealth 2016;4(1):e17 [FREE Full text] [doi: 10.2196/mhealth.5181] [Medline: 26872940]
36. Sheehan D, Lecrubier Y, Sheehan KH, Janavs J, Weiller E, Keskiner A, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. Eur Psychiatry 1997;12(5):232-241. [doi: 10.1016/s0924-9338(97)83297-x]
37. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
38. Boudreau DM, Lapham G, Johnson EA, Bobb JF, Matthews AG, McCormack J, et al. Documented opioid use disorder and its treatment in primary care patients across six U.S. health systems. J Subst Abuse Treat 2020 Mar;112S:41-48 [FREE Full text] [doi: 10.1016/j.jsat.2020.02.001] [Medline: 32220410]
39. Lapham GT, Lee AK, Caldeiro RM, Glass JE, Carrell DS, Richards JE, et al. Prevalence of behavioral health conditions across frequency of cannabis use among adult primary care patients in Washington State. J Gen Intern Med 2018 Nov;33(11):1833-1835 [FREE Full text] [doi: 10.1007/s11606-018-4558-8] [Medline: 29992423]
40. Lapham GT, Lee AK, Caldeiro RM, McCarty D, Browne KC, Walker DD, et al. Frequency of cannabis use among primary care patients in Washington State. J Am Board Fam Med 2017;30(6):795-805 [FREE Full text] [doi: 10.3122/jabfm.2017.06.170062] [Medline: 29180554]
41. Richards JE, Bobb JF, Lee AK, Lapham GT, Williams EC, Glass JE, et al. Integration of screening, assessment, and treatment for cannabis and other drug use disorders in primary care: an evaluation in three pilot sites. Drug Alcohol Depend 2019 Aug 01;201:134-141 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.04.015] [Medline: 31212213]
42. Sayre M, Lapham GT, Lee AK, Oliver M, Bobb JF, Caldeiro RM, et al. Routine assessment of symptoms of substance use disorders in primary care: prevalence and severity of reported symptoms. J Gen Intern Med 2020 Apr;35(4):1111-1119. [doi: 10.1007/s11606-020-05650-3] [Medline: 31974903]
43. Edwards P, Cooper R, Roberts I, Frost C. Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires. J Epidemiol Community Health 2005 Nov;59(11):987-999 [FREE Full text] [doi: 10.1136/jech.2005.034397] [Medline: 16234429]
44. Doody MM, Sigurdson AS, Kampa D, Chimes K, Alexander BH, Ron E, et al. Randomized trial of financial incentives and delivery methods for improving response to a mailed questionnaire. Am J Epidemiol 2003 Apr 01;157(7):643-651. [doi: 10.1093/aje/kwg033] [Medline: 12672684]
45. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
46. Hasin DS, Greenstein E, Aivadyan C, Stohl M, Aharonovich E, Saha T, et al. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): procedural validity of substance use disorders modules through clinical re-appraisal in a general population sample. Drug Alcohol Depend 2015 Mar 01;148:40-46 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.011] [Medline: 25604321]
47. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 2016 Nov 23;34(4):457-468. [doi: 10.1177/001872089203400407]
48. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
49. Adlin T, Pruitt J. The Essential Persona Lifecycle: Your Guide to Building and Using Personas. Burlington, Massachusetts, United States: Morgan Kaufmann; 2010:1-10.
50. Truong K, Hayes G, Abowd G. Storyboarding: an empirical determination of best practices and effective guidelines. In: Proceedings of the 6th conference on Designing Interactive systems. 2006 Presented at: DIS06: Designing Interactive Systems 2006; June 26-28, 2006; University Park PA USA p. 12-21. [doi: 10.1145/1142405.1142410]
51. Mares M, Gustafson DH, Glass JE, Quanbeck A, McDowell H, McTavish F, et al. Implementing an mHealth system for substance use disorders in primary care: a mixed methods study of clinicians' initial expectations and first year experiences.

XSL•FO
RenderX

BMC Med Inform Decis Mak 2016 Sep 29;16(1):126 [FREE Full text] [doi: 10.1186/s12911-016-0365-5] [Medline: 27687632]
52. Lapham GT, Hawkins EJ, Chavez LJ, Achtmeyer CE, Williams EC, Thomas RM, et al. Feedback from recently returned veterans on an anonymous web-based brief alcohol intervention. Addict Sci Clin Pract 2012 Aug 28;7:17 [FREE Full text] [doi: 10.1186/1940-0640-7-17] [Medline: 23186354]
53. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
54. Jochems EC, Mulder CL, Duivenvoorden HJ, van der Feltz-Cornelis CM, van Dam A. Measures of motivation for psychiatric treatment based on self-determination theory: psychometric properties in Dutch psychiatric outpatients. Assessment 2014 Aug 02;21(4):494-510. [doi: 10.1177/1073191113517928] [Medline: 24391079]
55. Ryan RM, Plant RW, O'Malley S. Initial motivations for alcohol treatment: relations with patient characteristics, treatment involvement, and dropout. Addict Behav 1995 May;20(3):279-297. [doi: 10.1016/0306-4603(94)00072-7]
56. Urbanoski KA, Wild TC. Assessing self-determined motivation for addiction treatment: validity of the Treatment Entry Questionnaire. J Subst Abuse Treat 2012 Jul;43(1):70-79. [doi: 10.1016/j.jsat.2011.10.025] [Medline: 22138201]
57. Cloud R, Kingree J. Concerns about dose and underutilization of twelve-step programs: models, scales, and theory that inform treatment planning. Recent Dev Alcohol 2008;18:283-301. [doi: 10.1007/978-0-387-77725-2_16] [Medline: 19115775]
58. Wild TC, Cunningham JA, Ryan RM. Social pressure, coercion, and client engagement at treatment entry: a self-determination theory perspective. Addict Behav 2006 Oct;31(10):1858-1872. [doi: 10.1016/j.addbeh.2006.01.002] [Medline: 16480834]
59. Wild T, Newton-Taylor B, Alletto R. Perceived coercion among clients entering substance abuse treatment: structural and psychological determinants. Addict Behav 1998 Jan;23(1):81-95. [doi: 10.1016/s0306-4603(97)00034-8]
60. Bauer AM, Rue T, Keppel GA, Cole AM, Baldwin L, Katon W. Use of mobile health (mHealth) tools by primary care patients in the WWAMI region Practice and Research Network (WPRN). J Am Board Fam Med 2014;27(6):780-788 [FREE Full text] [doi: 10.3122/jabfm.2014.06.140108] [Medline: 25381075]
61. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
62. McClure JB, Anderson ML, Bradley K, An LC, Catz SL. Evaluating an adaptive and interactive mhealth smoking cessation and medication adherence program: a randomized pilot feasibility study. JMIR Mhealth Uhealth 2016 Aug 03;4(3):e94 [FREE Full text] [doi: 10.2196/mhealth.6002] [Medline: 27489247]
63. Andréasson S, Danielsson A, Wallhed-Finn S. Preferences regarding treatment for alcohol problems. Alcohol Alcohol 2013;48(6):694-699. [doi: 10.1093/alcalc/agt067] [Medline: 23842842]
64. Dedoose Version 8.0.35, web application for managing, analyzing, and presenting qualitative and mixed method research data. SocioCultural Research Consultants, Los Angeles, CA2018. URL: https://www.dedoose.com/ [accessed 2021-05-24]
65. King N. Using templates in the thematic analysis of text. In: Cassell C, Symon G, editors. Essential Guide to Qualitative Methods in Organizational Research. Thousand Oaks, California, United States: SAGE Publications Ltd; 2004:256-270.
66. Creswell J. Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 3rd Ed. Thousand Oaks, CA: SAGE Publications; 2013:1-448.
67. Iversen RR. Book Review: Padgett, D. K. (2008). Qualitative Methods in Social Work Research (2nd ed.). Thousand Oaks, CA: Sage. 304 pp., $49.95 (paperback), ISBN 978-1-4129-5193-7; $84.95 (hardbound), ISBN 978-1-4129-5192-0. Res Soc Work Pract 2009 Dec 03;20(1):125-126. [doi: 10.1177/1049731509347859]
68. Beyer H, Holtzblatt K. Contextual Design. San Francisco, CA: Morgan Kaufman; 1998:1-496.
69. Wallhed FS, Bakshi A, Andréasson S. Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. Subst Use Misuse 2014 May;49(6):762-769. [doi: 10.3109/10826084.2014.891616] [Medline: 24601784]
70. Barry CL, Epstein AJ, Fiellin DA, Fraenkel L, Busch SH. Estimating demand for primary care-based treatment for substance and alcohol use disorders. Addiction 2016 Aug;111(8):1376-1384 [FREE Full text] [doi: 10.1111/add.13364] [Medline: 26899802]
71. Christensen DR, Landes RD, Jackson L, Marsch LA, Mancino MJ, Chopra MP, et al. Adding an internet-delivered treatment to an efficacious treatment package for opioid dependence. J Consult Clin Psychol 2014 Dec;82(6):964-972 [FREE Full text] [doi: 10.1037/a0037496] [Medline: 25090043]
72. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
73. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, REEACT collaborative. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. Br J Psychiatry 2017 May;210(5):362-367. [doi: 10.1192/bjp.bp.116.192435] [Medline: 28254959]

XSL•FO
RenderX

74. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020 Sep 01;77(9):906-914. [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
75. Muñoz RF. The efficiency model of support and the creation of digital apothecaries. Clin Psychol: Sci Pract 2017 Mar;24(1):46-49. [doi: 10.1111/cpsp.12174]
76. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. J Med Internet Res 2010 Dec 19;12(5):e48 [FREE Full text] [doi: 10.2196/jmir.1370] [Medline: 21169164]
77. Perron BE, Mowbray OP, Glass JE, Delva J, Vaughn MG, Howard MO. Differences in service utilization and barriers among Blacks, Hispanics, and Whites with drug use disorders. Subst Abuse Treat Prev Policy 2009 Mar 13;4(1):-. [doi: 10.1186/1747-597x-4-3]
78. Copeland J. A qualitative study of barriers to formal treatment among women who self-managed change in addictive behaviours. J Subst Abuse Treat 1997;14(2):183-190. [doi: 10.1016/s0740-5472(96)00108-0] [Medline: 9258863]
79. Cunningham JA, Sobell LC, Sobell MB, Agrawal S, Toneatto T. Barriers to treatment: why alcohol and drug abusers delay or never seek treatment. Addict Behav 1993 May;18(3):347-353. [doi: 10.1016/0306-4603(93)90036-9]
80. Rapp RC, Xu J, Carr CA, Lane DT, Wang J, Carlson R. Treatment barriers identified by substance abusers assessed at a centralized intake unit. J Subst Abuse Treat 2006 Apr;30(3):227-235 [FREE Full text] [doi: 10.1016/j.jsat.2006.01.002] [Medline: 16616167]
81. Sexton RL, Carlson RG, Leukefeld CG, Booth BM. Barriers to formal drug abuse treatment in the rural south: a preliminary ethnographic assessment. J Psychoactive Drugs 2008 Jun;40(2):121-129. [doi: 10.1080/02791072.2008.10400621] [Medline: 18720660]
82. van der Pol P, Liebregts N, de Graaf R, Korf DJ, van den Brink W, van Laar M. Facilitators and barriers in treatment seeking for cannabis dependence. Drug Alcohol Depend 2013 Dec 01;133(2):776-780. [doi: 10.1016/j.drugalcdep.2013.08.011] [Medline: 24035185]
83. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
84. Fang ML, Canham SL, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. Gerontologist 2019 Jan 09;59(1):1-15. [doi: 10.1093/geront/gny037] [Medline: 29750241]

Abbreviations

DUD: drug use disorder
PCP: primary care provider
PI: principal investigator

Edited by R Kukafka; submitted 24.11.20; peer-reviewed by G Hu, M Mueller; comments to author 22.01.21; revised version received 11.03.21; accepted 04.05.21; published 06.07.21.

Please cite as:
Glass JE, Matson TE, Lim C, Hartzler AL, Kimbel K, Lee AK, Beatty T, Parrish R, Caldeiro RM, Garza McWethy A, Curran GM, Bradley KA
Approaches for Implementing App-Based Digital Treatments for Drug Use Disorders Into Primary Care: A Qualitative, User-Centered Design Study of Patient Perspectives
J Med Internet Res 2021;23(7):e25866
URL: https://www.jmir.org/2021/7/e25866
doi:10.2196/25866
PMID:34255666

©Joseph E Glass, Theresa E Matson, Catherine Lim, Andrea L Hartzler, Kilian Kimbel, Amy K Lee, Tara Beatty, Rebecca Parrish, Ryan M Caldeiro, Angela Garza McWethy, Geoffrey M Curran, Katharine A Bradley. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Unstructured Supplementary Service Data System for Daily Tracking of Patient Samples and Diagnostic Results in a Diagnostic Network in Malawi: System Development and Field Trial

Daniel Killian[1*], BSc, MSc; Emma Gibson[1*], BSc, MSc; Mphatso Kachule[2], BEd, MSc; Kara Palamountain[3], BBA, MBA; Joseph Bitilinyu Bangoh[4], MSc; Sarang Deo[5], PhD; Jonas Oddur Jonasson[6], PhD

[1]Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, United States
[2]Riders 4 Health, Lilongwe, Malawi
[3]Kellogg School of Management, Northwestern University, Evanston, IL, United States
[4]Health Technical Services, Malawi Ministry of Health, Lilongwe, Malawi
[5]Indian School of Business, Hyderabad, India
[6]MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States
[*]these authors contributed equally

Corresponding Author:
Jonas Oddur Jonasson, PhD
MIT Sloan School of Management
Massachusetts Institute of Technology
100 Main Street
Cambridge, MA, 02139
United States
Phone: 1 617 715 5109
Email: joj@mit.edu

Abstract

Background: Diagnostics in many low- and middle-income countries are conducted through centralized laboratory networks. Samples are collected from patients at remote point-of-care health facilities, and diagnostic tests are performed at centralized laboratories. Sample transportation systems that deliver diagnostic samples and test results are crucial for timely diagnosis and treatment in such diagnostic networks. However, they often lack the timely and accurate data (eg, the quantity and location of samples prepared for collection) required for efficient operation.

Objective: This study aims to demonstrate the feasibility, adoption, and accuracy of a distributed data collection system that leverages basic mobile phone technology to gather reports on the quantity and location of patient samples and test results prepared for delivery in the diagnostic network of Malawi.

Methods: We designed a system that leverages unstructured supplementary service data (USSD) technology to enable health workers to submit daily reports describing the quantity of transportation-ready diagnostic samples and test results at specific health care facilities, free of charge with any mobile phone, and aggregate these data for sample transportation administrators. We then conducted a year-long field trial of this system in 51 health facilities serving 3 districts in Malawi. Between July 2019 and July 2020, the participants submitted daily reports containing the number of patient samples or test results designated for viral load, early infant diagnosis, and tuberculosis testing at each facility. We monitored daily participation and compared the submitted USSD reports with program data to assess system feasibility, adoption, and accuracy.

Results: The participating facilities submitted 37,771 reports over the duration of the field trial. Daily facility participation increased from an average of 50% (26/51) in the first 2 weeks of the trial to approximately 80% (41/51) by the midpoint of the trial and remained at or above 80% (41/51) until the conclusion of the trial. On average, more than 80% of the reports submitted by a facility for a specific type of sample matched the actual number of patient samples collected from that facility by a courier.

Conclusions: Our findings suggest that a USSD-based system is a feasible, adoptable, and accurate solution to the challenges of untimely, inaccurate, or incomplete data in diagnostic networks. Certain design characteristics of our system, such as the use

XSL•FO
RenderX

of USSD, and implementation characteristics, such as the supportive role of the field team, were necessary to ensure high participation and accuracy rates without any explicit financial incentives.

(J Med Internet Res 2021;23(7):e26582) doi:10.2196/26582

KEYWORDS

diagnostic networks; mobile phone; sample transportation; sub-Saharan Africa; data collection

Introduction

Background

Most populations in sub-Saharan Africa rely on rural health clinics that have poor infrastructure as their primary point of entry to a broader health care system. These health clinics are often not equipped to conduct diagnostic testing, but they refer patient samples to a relatively small number of centralized laboratories capable of diagnostic analysis [1]. The effectiveness of programs targeting active diseases in the region, such as HIV/AIDS, tuberculosis (TB), and malaria, is therefore closely related to the performance of sample transportation (ST) systems put in place to transport patient samples and test results across the difficult terrain separating health facilities and molecular laboratories (MLs) [2].

ST systems in many sub-Saharan African countries operate without accurate information regarding the quantity and location of patient samples and test results requiring transportation [3]. Consequently, these ST systems operate in *push* mode, where couriers visit facilities on fixed weekly or biweekly schedules [3,4]. A common result of this operating mode is *empty trips*: courier visits to facilities where nothing was delivered to, or transported from, the facility. For example, an analysis of archival 2017-2018 courier data in Malawi revealed that 30.08% (24,256/80,648) of the courier visits to clinics were empty trips. This not only results in inefficient use of limited resources but also contributes to delays in receiving critical test results [5], which in turn leads to poor health-seeking behavior among the population [6] and can contribute to increased mortality rates [7].

An alternative to these ST push systems would be a *pull* system in which couriers only visit facilities when patient samples or test results are ready for transport to, or from, that facility. Such a pull system would limit empty trips but would require a reliable system to track the number of patient samples and test results requiring transportation across the diagnostic network. We hypothesize that these logistics data—specifically, the location and quantity of patient samples or test results ready for transport—can be collected with a low-cost information-sharing system and can be used to create an ST system that is responsive to real-time needs.

In this study, we investigated the feasibility, adoption, and accuracy of a system that leverages a communication protocol that is standard on all mobile phones, known as unstructured supplementary service data (USSD) technology [8], to gather more timely and accurate information. To conduct this investigation, we designed and developed a USSD collection system (hereafter, *the USSD system*) to enable health care facilities in a diagnostic network to report daily sample volume data. We conducted a year-long field trial of the USSD system in Malawi from July 2019 to July 2020 to determine whether our system would enable the timely collection of information.

Intervention Development and Design

The 2 critical design questions we faced when developing our system were as follows:

1. What technology would health workers use to submit sample volume reports?
2. How would these reports be structured?

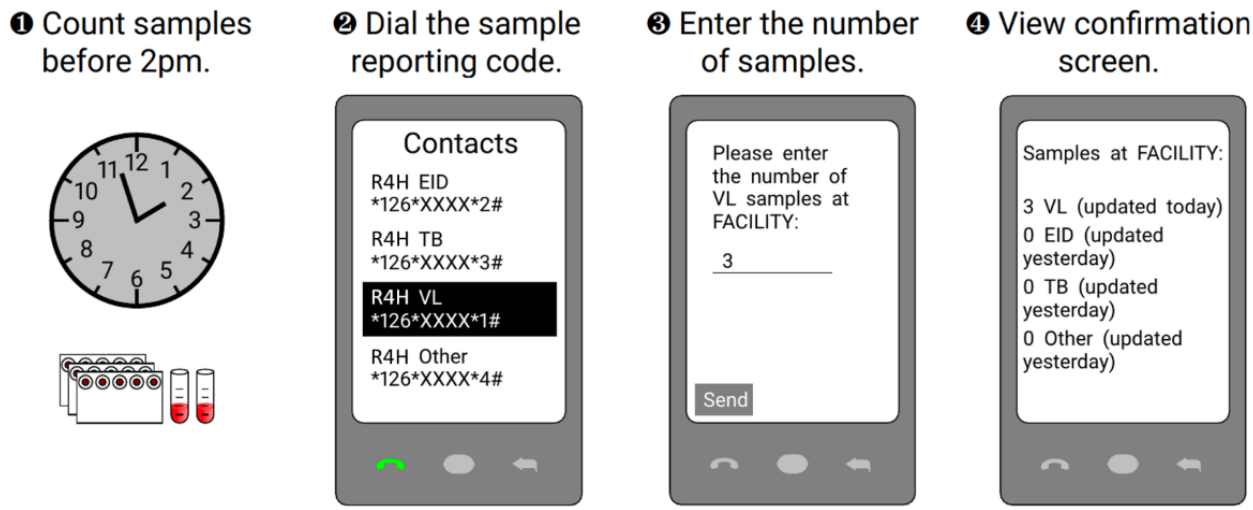
Given the positive findings regarding the feasibility of mobile health (mHealth) initiatives in low- and middle-income countries [9] and the rapid growth in mobile network coverage in sub-Saharan Africa over the past decade [10,11], we designed our system so that health workers could submit reports with a mobile phone. As less than half of the mobile phones in the region are smartphones [11], we based our sample volume data collection system on USSD technology—a mobile communication protocol accessible to both smartphones and less technologically advanced feature phones.

In a USSD system, users are provided a unique numeric access code that they can dial to access a structured menu of options. With an appropriate series of key presses using the mobile phone's keypad, users can navigate through the options menu and submit information, similar to the method that an SMS text messaging system uses to convey information through text. One advantage of USSD over SMS text messaging, especially in the context of data collection, is that the USSD's built-in menu system allows for structured responses, leading to built-in data validation [8]. This also reduces the amount of effort required to leverage the data systematically, which is crucial for managing the growing volume of health care data received through mobile phones [12].

In our USSD system, a specific USSD reporting code is assigned to every health facility operating within a diagnostic network, and each diagnostic test offered in the network is assigned a unique numeric designator. To report the number of patient samples for a specific test, a designated health worker at a designated facility can use any mobile phone to dial that facility's USSD reporting code and the numeric designator for the diagnostic test to connect to the USSD system. Once connected, they can report the number of patient samples through a text-based interface. Upon entering all the required information, the user receives a confirmation text message containing a summary of the submitted report. Refer to Figure 1 for further details.

XSL•FO
RenderX

Figure 1. Unstructured supplementary service data system reporting instructions. (1) A health worker counts the number of samples, by sample type, that are prepared for transport. (2) The health worker dials a USSD code specifically assigned for the health worker's clinic and the type of sample (EID, VL, TB, and other), which has been saved on the health worker's phone and is displayed on a poster in the health facility. (3) The health worker enters the number of samples currently awaiting transportation at the facility. (4) The health worker views a confirmation screen, which summarizes how many samples of each type have been reported for the facility. EID: early infant diagnosis; R4H: Riders for Health; TB: tuberculosis; USSD: unstructured supplementary service data; VL: viral load.



Methods

Study Setting

As of 2018, an estimated 9.75% (790,000/8,100,000) of the adults (aged 15-49 years) in Malawi were living with HIV [13], and the country's TB incidence rate was 181 per 100,000 people [14]. The Malawi Ministry of Health (MOH) operates a diagnostic network of approximately 700 widely distributed community health clinics, 27 centrally located district health offices, and 10 regionally aligned MLs. The structure of the Malawi diagnostic network is representative of the diagnostic networks of many other countries in the region [2].

Since 2016, the Malawi branch of the nonprofit organization Riders for Health International (R4H) has managed the transportation of viral load (VL) samples, early infant diagnosis (EID) tests, and TB samples from health clinics and district health offices to the MLs. R4H maintains a team of more than 70 motorcycle couriers who visit health facilities and laboratories according to fixed weekly schedules.

In collaboration with MOH and R4H, we identified 3 districts in Malawi to test the USSD system: Salima in the central region, Rumphi in the northern region, and Phalombe in the southern region. The diagnostic networks in these rural and semirural districts each contained between 15 and 18 facilities and relied upon one ML to conduct the diagnostic testing for that diagnostic network, making these districts a representative sample of R4H's typical ST operations in rural and semirural areas.

This field trial was approved by Malawi's National Health Sciences Research Council. An evaluation by the Massachusetts Institute of Technology Committee on the use of humans as experimental subjects determined that the trial did not constitute human subjects research, as defined in Federal Regulations 45CFR46.

System Implementation

In early 2019, we contracted with a local vendor in Malawi to develop the user interface and information technology infrastructure. The vendor also managed the daily operation of the system, which included storing incoming data, sending reminder messages when appropriate, and contracting with the cellular network providers to enable free provision of the USSD service to health workers.

In May 2019, we asked the facility in-charge from each of the participating health facilities to nominate 1 or 2 staff members with personal mobile phones to enter the data. In June 2019, we conducted three 2-hour training sessions (1 per study district) to train 150 health workers to use the USSD system. During the training sessions, we introduced the USSD system to the participants, taught the study participants how to access the system and submit reports using their personal mobile device, and provided reference posters and flyers that reminded the participants how to access the system for them to display at their facilities. A field team, consisting of a local field manager and 3 local research assistants, monitored the implementation and addressed technical and logistical challenges through regular communication with health workers, district laboratory technicians, and R4H couriers through phone calls and text messages.

The USSD system was officially launched in the study districts in July 2019. Health workers were asked to report the number of patient samples waiting to be transported at the end of each day. Facilities were expected to submit a report every day, even if they had not collected any samples or prepared any new samples for transportation. Health workers and field team members were sent a series of automated daily reminder messages to increase participation (Textbox 1).

XSL•FO
RenderX

Textbox 1. The daily schedule of reminder notifications.

Daily Reminder Notifications: Notification Delivery Time and Content

- 8 AM: Health workers at each facility are sent a message notifying them whether or not a courier will visit their facility later that day as well as a reminder to report sample volumes.
- Noon: Health workers at each facility are sent a second reminder to report sample volumes.
- 1:30 PM: Members of the field team are sent a summary of the facilities for which a report has or has not been submitted.
- 2:15 PM: Health workers at facilities missing all or part of a complete daily report are reminded to report sample volumes.
- 3 PM: Members of the field team are sent an updated summary of the facilities for which a report has or has not been submitted as well as a comparison of each facility's current report with its previous report (because an unusual increase or decrease from the previous report may indicate that the facility is reporting incorrectly).
- 4:15 PM: Members of the field team are sent a notification informing them whether couriers have submitted reports to the courier database. Reports submitted to the courier database are summarized by facility, compared with that facility's most recent unstructured supplementary service data (USSD) report, and sent to members of the field team to assess facility reporting accuracy.
- 7 PM: Members of the field team are sent an updated list of the couriers who have or have not submitted reports to the courier database. A summary of reports submitted to the courier database by the facility are recompiled to capture any updates, recompared with that facility's most recent USSD report, and sent to members of the field team to assess facility reporting accuracy.

On the basis of the notifications received from the health facilities, the field team sought out any unusual participation patterns such as intermittent, erratic, or extended periods of no participation. Upon detection of unusual participation patterns, the field team was authorized to directly address these patterns with the participant. In situations where an ordinarily reliable participant simply forgot to report or health workers in the same facility failed to properly delegate reporting responsibilities, the field team could directly contact the designated staff member at the facility. The field team could then remind the designated staff member to submit their daily report or to delegate reporting responsibilities to a different health worker when the primary contact was not at the facility. If the field team identified intermittent network coverage as the cause of a missed report, the field team could delegate reporting responsibility to someone with a network connection. In addition, the field team could also ask the courier to hand-deliver a message to the responsible health worker, to identify someone else at the facility to accept reporting responsibilities, or to request escalation to a higher authority at the nonreporting facility in email notifications regarding their facility's participation.

The field team also monitored the accuracy of reports and intervened directly with the participants if they observed a pattern of low-accuracy reports. As in the case of poor participation, the field team could use a combination of phone calls, text messages, and hand-delivered messages to identify the root causes of data inaccuracies and address them. The preferred approach for improving reporting accuracy was to provide additional instructions to the noncompliant participant. If a training update failed to address the situation, more drastic measures (eg, requesting transfer of reporting to another staff member) could be adopted.

Evaluation Framework

As part of the USSD system implementation plan, we elected to evaluate the feasibility, adoption, and accuracy of the USSD system using relevant descriptive statistics. Feasibility and adoption are common evaluation domains in the intervention assessment literature [15,16]. Accuracy, although not a common evaluation domain regarding health interventions, is relevant in the context of mobile device–based data collection systems [17]. Textbox 2 lists the guiding questions and associated metrics for assessing the system's performance within the 3 domains. The feasibility of the USSD system depended on whether each facility had access to the technology required to engage with it. Therefore, we identified the number of facilities employing someone with a mobile device who was willing to participate in the study and the number of facilities in the field trial districts receiving service from a wireless network provider.

To assess the adoption of the USSD system, we monitored specific participation-related metrics: percentage of facilities reporting by day, individual facility participation over the course of the field trial, and the longest period that each facility went without participating.

We determined the accuracy of the USSD system by comparing the submitted USSD reports with program data. A data report was deemed accurate if the reported number of patient samples of a given type ready for delivery and the actual number of patient samples ready for delivery, as determined by the couriers, were identical.

XSL•FO
RenderX

Textbox 2. Evaluation framework.

Domains, Guiding Questions, and Metrics of the Evaluation Framework

Feasibility

- Do facilities have access to mobile devices?
  - The fraction of facilities for which a personal mobile phone was registered.
- Do facilities receive a mobile network signal?
  - The fraction of facilities where insufficient network connection never prevented that facility from submitting a report.
  - The fraction of facilities where staff members at that facility submitted a daily report for at least 7 consecutive days.

Adoption

- Are facilities participating?
  - The fraction of facilities that reported or failed to report each day by sample type.
  - The fraction of total reporting days over the trial period when each facility reported or failed to report.
  - The largest number of consecutive days that a facility failed to report.
- Which operational factors influenced facility participation?
  - The fraction of facilities where an insufficient understanding of the unstructured supplementary service data (USSD) system on the part of health workers prevented USSD participation.
  - The fraction of facilities where hardware limitations prevented USSD participation.
  - The fraction of facilities where health worker workload prevented USSD participation.
  - The fraction of facilities where health worker absences prevented USSD participation.
  - The fraction of facilities where health worker forgetfulness prevented USSD participation.

Accuracy

- How accurate are the data reported by participating facilities?
  - The average and variance of the difference between reported and actual sample volumes.

Data

We used data from 4 distinct sources to calculate the metrics listed in Textbox 2: the USSD system database, the courier database, a survey administered to members of the field team, and the attendance roster from the USSD system training sessions.

Every data report submitted through the USSD system during the field trial was archived in the USSD system database. This database included the facility name, date and time, the user's mobile number, sample type, and number of patient samples reported by the user for every data report.

The courier database contained sample-specific information submitted by the couriers upon completion of the couriers' daily routes to a data collection system operated by R4H. The sample data captured in the courier database consisted of the sample's identification code, the name of the facility the sample originated from, the date of sample collection, and the date of sample pick-up from the originating facility.

Upon completion of the study, we administered a survey to the research assistants to assess the barriers to system participation. For each of the 51 facilities included in the field trial, the local research assistants were asked the questions presented in Textbox 3.

We compiled a master attendance roster by combining the individual attendance rosters recorded at each of the 3 USSD system training sessions. These rosters included the name of each training participant, the facility the participant represented, the participant's staff position, and the participant's contact information.

XSL•FO
RenderX

Textbox 3. Questions asked in the survey to research assistants.

Survey Questions Answered by Each Research Assistant

- For each facility in your district, estimate the number of times poor network reception prevented the facility from reporting.
- Rate the effectiveness of the following techniques on participation by facilities in your district:
  - SMS text messages
  - Individual messages via a popular internet messaging platform
  - Phone calls
  - Asking a courier to deliver a message
  - In-person facility visits
  - Group messages via a popular internet messaging platform

Analysis

To calculate the percentage of facilities for which a personal mobile phone was registered, we reviewed the master training attendance roster. Attendance at a training event by an employee from a given facility indicated that the employee owned a mobile phone and was willing to use their device to submit data reports to the USSD system. To calculate the percentage of facilities with a sufficient network connection, we summarized the survey responses regarding the frequency with which poor network connectivity prevented each facility from participating.

To determine the number of facilities for which a USSD report was submitted for at least 7 consecutive days, we analyzed the facility name and date of every report submitted to the USSD system database. Aggregation and summarization of data in the USSD system database also allowed us to measure all 3 metrics associated with the first guiding question in the adoption domain (Textbox 2).

We calculated the accuracy of the reported data by comparing the reported data in the USSD system database with the courier reports in the courier database, which captured the number of patient samples collected from the health care facilities.

All data analyses were conducted using the R (v 4.0.0; R Foundation for Statistical Computing) language and RStudio Desktop (v 1.2.5042). Reported *P* values were calculated using 1-sided nonparametric Mann-Whitney tests (unless otherwise noted).

Results

Descriptive Statistics

Over the study period (July 2019 to July 2020), the participating facilities submitted 37,771 reports to the USSD system, accounting for 48,852 patient samples. Most of these patient samples (40,952/48,852, 83.83%) were VL samples, whereas 6.1% (2979/48,852) were EID samples, 5.85% (2859/48,852) were TB samples, and 4.21% (2056/48,852) were classified as *Other*. Of the samples reported, 43.71% (21,355/48,852) originated in Phalombe, 35.12% (17,155/48,852) originated in Salima, and 21.17% (10,342/48,852) originated in Rumphi. The table included in Multimedia Appendix 1 contains sample volume statistics by district.

Feasibility

All participating facilities employed at least one individual willing to submit reports to the USSD system with a personal mobile device. The research assistants reported that an insufficient network connection never prevented 47% (24/51) of the participating facilities from submitting a report to the USSD system, caused occasional submission problems in 24% (12/51) of the facilities, and caused frequent problems in 29% (15/51) of the facilities. Our analysis of the USSD database data also revealed that each facility had at least one 7-day period during which the facility submitted a report every day.

Adoption

Figure 2 illustrates the daily sample reporting rates and the 7-day moving average of daily sample reporting rates for the 3 patient sample types between July 2019 and July 2020. At the beginning of the study, only 10% (5/51)-20% (10/51) of the facilities participated each day. However, after 3 weeks, the daily participation rates increased and remained between 53% (27/51) and 98% (50/51) for VL samples, between 51% (26/51) and 98% (50/51) for EID samples, and between 43% (22/51) and 96% (49/51) for TB samples. Between August 2019 and January 2020, the 7-day moving average participation rate increased gradually for VL (from 63% to 87%), EID (from 60% to 85%), and TB (from 54% to 79%) samples, with notable but temporary declines during the second halves of November and December. For the final 5 months of the trial (February 2020 to July 2020), the average participation rate remained at or above 75% (38/51) for all 3 sample types.

The distribution of the facility participation rates across all districts and by individual districts, where the facility participation rate is calculated as the percentage of days out of the total possible reporting days for which a facility reported, is shown in Figure 3. On average, the facilities provided a report 78.9% (198/251 possible reporting days) of the time (σ=32.6 days). The median number of days a facility reported was 81.3% (204/251) days, with a range from 48.2% (121/251) days to 97.6% (245/251) days. The facilities in Phalombe reported less frequently than those in Salima (*P*=.003) and Rumphi (*P*=.01) on average.

XSL•FO
RenderX

Figure 2. Facility participation by sample type, shown as daily participation percentages and as a 7-day moving average. EID: early infant diagnosis; TB: tuberculosis; VL: viral load.

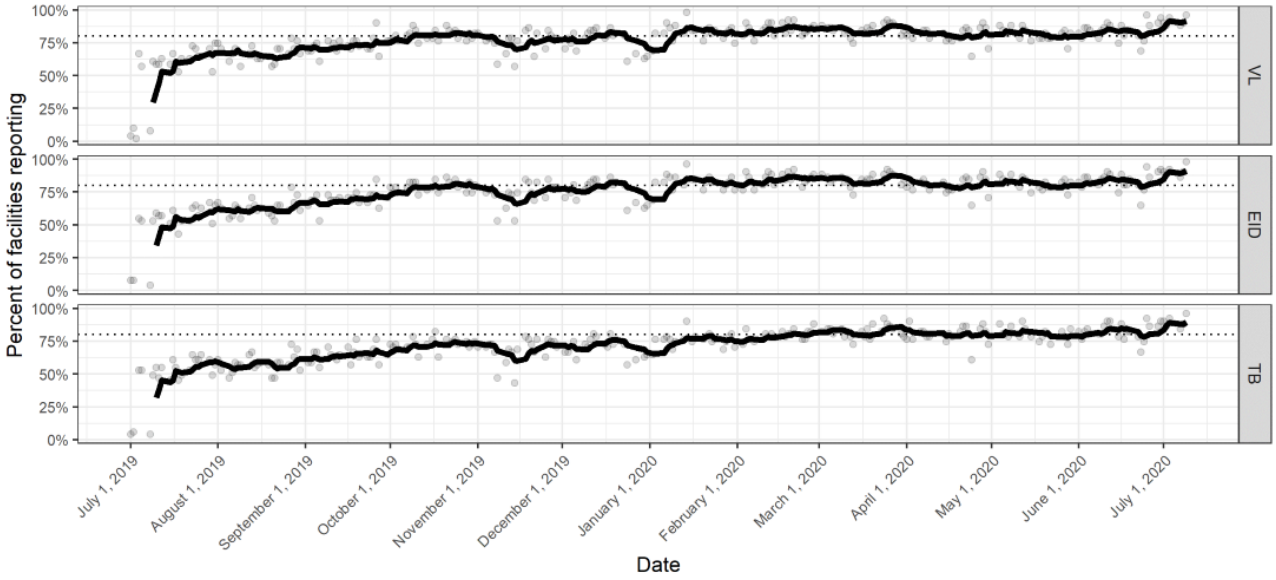
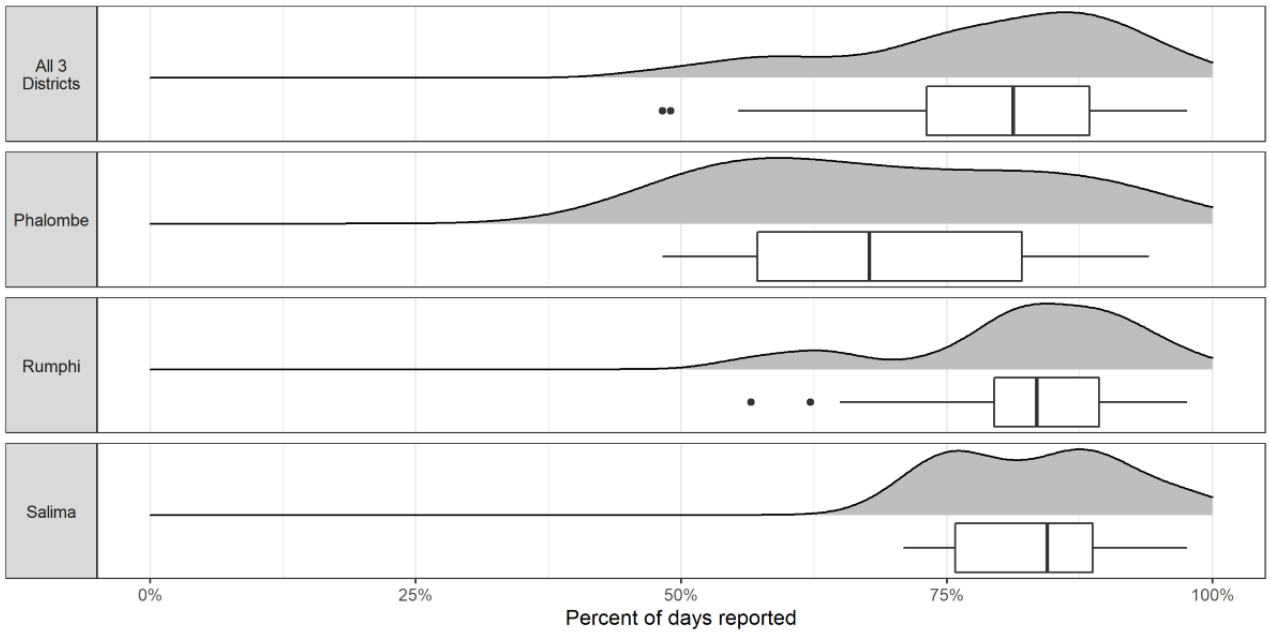


Figure 3. Distribution of facility reporting frequency.



Of the 51 health facilities, 45% (23/51) never went more than 1 business week (5 days) without submitting a report, and the longest any facility went without providing a report was 30 days. On average, the longest a facility went without submitting a report was 8.65 days (σ=6.33 days).

Table 1 shows the frequency of participation challenges faced by health facilities according to the 3 local research assistants. Each cell shows the number and percentage of facilities reported to experience a specific concern to the given extent. Recurring compliance issues were most often due to poor network reception, whereas the occasional noncompliance issue was most likely due to forgetfulness on the part of the health worker. Staff absence at the participating facilities also caused reporting issues at most (31/51, 61% facilities) of the participating facilities. These research assistant survey results also suggest that health workers had an adequate understanding of the system; poor staff training did not cause any reporting problems in more than 80% (41/51) of the facilities.

XSL•FO
RenderX

Table 1. Results regarding the causes of reporting issues at each of the 51 participating facilities (according to the 3 research assistants; n=51).

Reason for no report	Facilities, n (%)		
	No problems	Occasional problems	Frequent problems
Insufficient mobile network reception	24 (47)	12 (24)	15 (29)
Phone issues (eg, low battery or broken phone)	29 (57)	18 (35)	4 (8)
Staff are absent	20 (39)	21 (41)	10 (20)
Staff are too busy	26 (59)	15 (29)	6 (12)
Staff do not understand how to use the system	43 (84)	8 (16)	0 (0)
Staff forgot to report	4 (8)	39 (76)	8 (16)

Accuracy

Figure 4 illustrates the daily percentage accuracy of reports for each sample type and its 7-day moving average. The daily accuracy of VL reports slowly improved over the first 2 months of the field trial and settled at approximately 80% for the remainder of the trial. Unlike the daily accuracy of VL reports, that of EID and TB reports did not change substantially throughout the trial, with the daily accuracy of VL reports exhibiting greater variance (σ=7.23) than both EID daily reporting accuracy (σ=5.25; *P*<.001 according to Levene test) and TB daily reporting accuracy (σ=5.38; *P*<.001 according to Levene test). On average, 81% of the daily VL reports, 89.2% of the daily EID reports, and 88.2% of the daily TB reports were accurate.

The distribution of data accuracy by sample type across facilities is shown in Figure 5. The accuracy of EID reports exhibited the least variation (σ=0.01), followed by VL reports (σ=0.11) and TB reports (σ=0.14). The median facility reporting accuracy for VL was 82%, which was lower than that for EID (91%; *P*=.001 according to a paired Mann-Whitney test) and TB (91%; *P*=.001; paired Mann-Whitney test). For each sample type, more than half of the facilities submitted accurate reports on more than 80% of the days in the field trial.

Figure 4. The number of accurate reports by sample type, shown as daily accuracy percentages and as a 7-day moving average. EID: early infant diagnosis; TB: tuberculosis; VL: viral load.

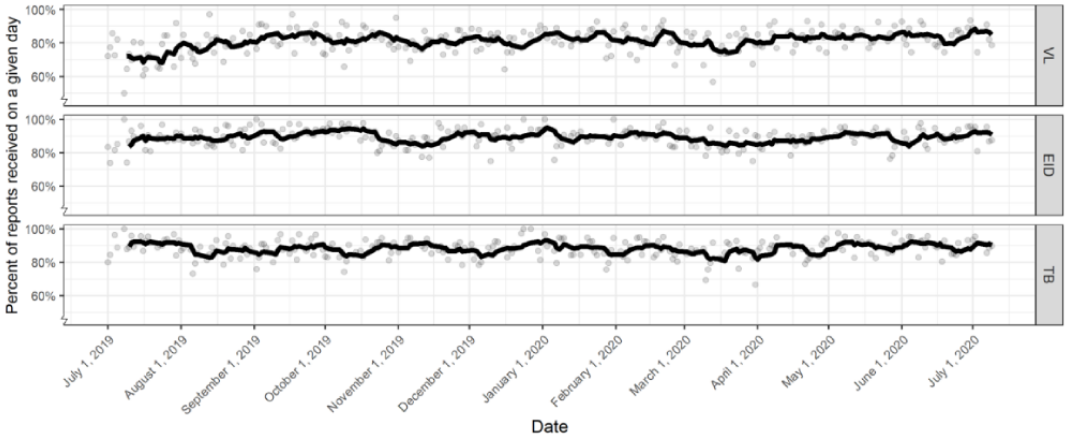
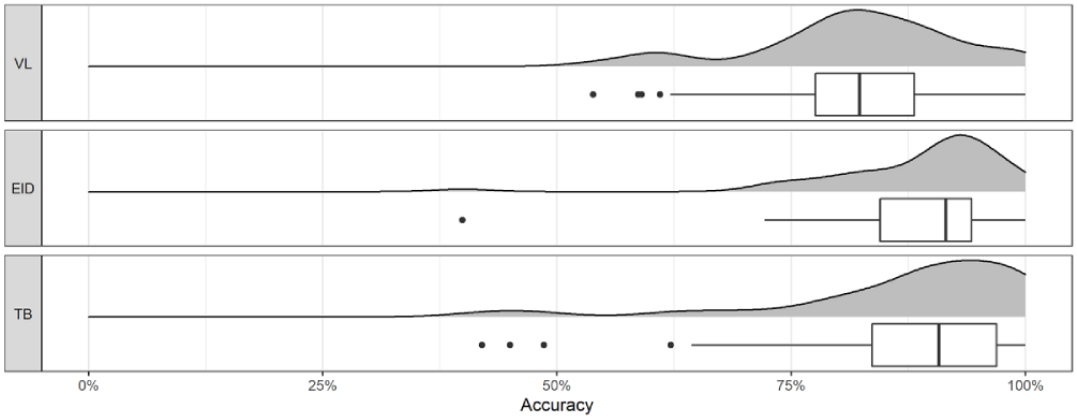


Figure 5. Distribution of facility reporting accuracy by sample type. EID: early infant diagnosis; TB: tuberculosis; VL: viral load.



XSL•FO
RenderX

Discussion

Principal Findings

General Findings

We designed a system whereby staff members in geographically dispersed health care facilities could report, through any mobile device, the number of patient samples prepared for delivery to a diagnostic test facility. Between July 2019 and July 2020, we conducted a field trial of the system in 3 districts in Malawi to assess the system's feasibility, adoption, and accuracy. The results from our field trial suggest that the USSD system is a feasible, adoptable, and accurate tool for assembling accurate daily reports on the quantity and location of transportation-ready patient samples and diagnostic test results.

Feasibility

The feasibility of the USSD system is driven by the ease with which facilities can access it for report submission with a mobile network connection and a mobile device. Although mobile network coverage varies by country, our findings with respect to the number of facilities able to submit a report through the system illustrate the potential of using mHealth systems to link rural health facilities to central operations managers in sub-Saharan Africa, especially as mobile network coverage continues to improve across the region [11]. We also found at least one person at each facility with a mobile phone—a significant result because less than half of the population in the region owns a mobile phone [18]. Although prior work in the region regarding mHealth initiatives among the general population has, at times, found poor eligibility rates among potential participants driven by low mobile phone ownership [19,20], our findings suggest that mHealth efforts requiring ownership among health workers may be more feasible than those requiring ownership among the general population [21]. In addition, it is likely that the use of USSD in a region where smartphones constitute less than 40% of all mobile phones with lower-tier connections (ie, 1G/2G as against 4G/5G) also contributed to system feasibility [11]. The use of health workers' personal phones avoided the additional cost of deploying new devices in the field and leveraged their familiarity with their devices to enhance system usability and transfer maintenance responsibilities to them [22].

Adoption

The adoption of the USSD system improved consistently over the course of the study and peaked near 90% toward the conclusion of the study, with the exception of small and temporary declines coinciding with personnel transitions (ie, staff reallocations and annual training sessions) and holiday seasons. These findings are comparable with results from similar mHealth studies conducted in the region [21,22], despite our study requiring more frequent reporting than other studies and doing so without monetary compensation for the participants. On the basis of the calculated descriptive statistics, we attribute the wide adoption of the USSD system to close collaboration with MOH representatives. This collaboration secured the support of senior government officials who encouraged participation by health workers at the facilities. In addition, this collaboration improved the chances of the system's design complementing health workers' existing responsibilities rather than adding to them, which is known to increase the likelihood of system adoption [19,21,23]. The efforts of the field team members in their role as real-time participation monitors and problem solvers likely also influenced the observed participation rate.

Accuracy

The existence of the courier database and our ability to access that data played a significant role in ensuring the high accuracy of the records (>80%). In contrast with prior mHealth initiatives [22,24], the courier database allowed us to assess the accuracy of every report submitted through our system with minimal delay and to provide feedback, via the field team, to correct improper reporting behavior daily. Providing timely and relevant feedback to a health worker regarding their reporting behavior likely contributed to the overall reporting accuracy achieved in the field trial [25].

Strengths and Limitations

The scope of our study is limited to establishing the feasibility, adoption, and accuracy of a USSD system for collecting information on the quantity and location of patient samples and test results prepared for delivery in the diagnostic network. Therefore, the impact of making these data available on subsequent operational decisions or on patient care remains undetermined. This information is expected to be useful for avoiding unnecessary health facility visits, but there are many ways to incorporate the information provided by the USSD system into an operational courier routing system aimed at reducing empty trips and delays. The rigorous quantification of the effect of the USSD system on ST operations is beyond the scope of this study. However, a recent report by Gibson et al [26] indicated that this information, along with a sophisticated routing system, can reduce empty trips by at least 55%.

In addition, the results presented in Table 1 regarding the operational factors affecting facility participation are based on data collected indirectly through a survey administered to the 3 research assistants in the field. Ideally, these data should have been collected directly from each facility because this would have provided more granular information about the operational drivers for participation. However, collecting operational data on a daily basis was beyond the scope of this effort, and the constant rotation of health workers into and out of the facilities over the course of the study made it infeasible to conduct an end-of-study survey at each facility. We believe that surveying the local research assistants was the next best solution because they were in regular contact with multiple health workers from each facility and were aware of their experiences with the USSD system. Regardless of this limitation, the main objective of this study was to assess the feasibility, adoption, and accuracy of collecting information using the USSD system, all of which can be evaluated using primary data sources.

A notable strength of our study is that it demonstrates a novel use of mHealth technology to significantly improve information sharing in diagnostic networks, which have a similar structure in many low- and middle-income countries. Previous mHealth

XSL•FO
RenderX

studies have investigated how mHealth technology can improve health care delivery through the wide dissemination of health-related information [9], providing patient-specific reminders and/or results to patients [27,28], connecting health care providers at different levels in the health care network [4,29], and monitoring medical supply stock levels [21,22,29], among other applications [19,30]. To the best of the authors' knowledge, this study represents the first application of mHealth technology to track the location of samples and results in a large-scale diagnostic network. Additional strengths of this study include the fact that field implementation was sustained for an entire year, and that the structure of the system allowed us to determine the accuracy of every submitted report.

Scalability

The USSD system is extremely scalable from a technological perspective because there is no requirement to purchase a specific mobile phone, mobile phone airtime, or any other system-specific technology. Expanding the system to operate with new facilities and/or new diagnostic tests simply requires assigning USSD identification codes and training new users, which itself is not very onerous because of the familiarity of mobile users in the region with USSD technology through other apps [8]. The scalability of the USSD system is further enhanced because it does not require the purchase of new hardware or software. The system operates with a mobile network signal that is increasingly available across most of the region and uses technology universally embedded in all mobile devices.

As explained earlier, the USSD system was designed to minimize impact on health workers' workload, which should positively affect adoption in other health facilities without disrupting their routines [31,32]. Furthermore, health workers who currently use the USSD system can share their experiences in training sessions for new facilities, thereby further speeding up adoption. However, scalability may be adversely affected by continued reliance on field teams for data monitoring and supervision. As we work on developing scale-up plans in collaboration with R4H and MOH, we believe that incorporating these tasks into the roles of senior personnel within the ST systems, such as the regional ST coordinators who oversee ST operations within the districts, can help overcome this problem and facilitate scalability.

Conclusions

Malawi's diagnostic network, both in terms of the network's structure and challenges, is representative of many diagnostic networks being operated in sub-Saharan Africa [2]. The descriptive results of our study suggest that a USSD-based system is a feasible, adoptable, and accurate solution to the challenges of untimely, inaccurate, or incomplete data present in these diagnostic networks. The scalability of the USSD system, along with the promising results of our study, suggests that system implementation at the national level in many sub-Saharan nations is feasible and worthwhile.

Acknowledgments

This study was supported by Grand Challenges Award #OPP1182217 from the Bill and Melinda Gates Foundation. The research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number U54EB027049. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

EG, JOJ, SD, KP, and MK designed the study. MK and JBB oversaw the field trial implementation. EG, JOJ, and SD conceptualized the manuscript, guided the analysis, and supported the writing of multiple drafts. DK conducted the literature review and analysis and conceptualized and drafted the manuscript. EG conducted the analysis and supported the design of the qualitative data collection tools and methodology. All authors reviewed multiple drafts of the manuscript and read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary statistics regarding the number of samples, by type, district, and facility, included in the unstructured supplementary service data system database from July 2019 to July 2020.
[DOCX File , 27 KB - jmir_v23i7e26582_app1.docx]

References

1. Lankowski AJ, Siedner MJ, Bangsberg DR, Tsai AC. Impact of geographic and transportation-related barriers on HIV outcomes in sub-Saharan Africa: a systematic review. AIDS Behav 2014 Jul 23;18(7):1199-1223 [FREE Full text] [doi: 10.1007/s10461-014-0729-8] [Medline: 24563115]
2. Schöpperle A. Analysis of challenges of medical supply chains in sub-Saharan Africa regarding inventory management and transport and distribution. Project Thesis, University of Westminster, London. 2013. URL: https://iaphl.org/wp-content/uploads/2016/05/Medical-Supply-Chain-Challenges.Masterthesis.ASchoepperle.pdf [accessed 2021-05-25]

XSL•FO
RenderX

3. Kiyaga C, Sendagire H, Joseph E, McConnell I, Grosz J, Narayan V, et al. Uganda's new national laboratory sample transport system: a successful model for improving access to diagnostic services for Early Infant HIV Diagnosis and other programs. PLoS One 2013 Nov 13;8(11):e78609 [FREE Full text] [doi: 10.1371/journal.pone.0078609] [Medline: 24236026]
4. Binagwaho A, Mugwaneza P, Irakoze AA, Nsanzimana S, Agbonyitor M, Nutt CT, et al. Scaling up early infant diagnosis of HIV in Rwanda, 2008-2010. J Public Health Policy 2013 Jan 29;34(1):2-16. [doi: 10.1057/jphp.2012.62] [Medline: 23191941]
5. Jónasson JO, Deo S, Gallien J. Improving HIV early infant diagnosis supply chains in sub-Saharan Africa: models and application to Mozambique. Oper Res 2017 Dec;65(6):1479-1493. [doi: 10.1287/opre.2017.1646]
6. Ndarukwa P, Chimbari MJ, Sibanda E. Assessment of levels of asthma control among adult patients with asthma at Chitungwiza Central Hospital, Zimbabwe. Allergy Asthma Clin Immunol 2020 Feb 4;16(1):10 [FREE Full text] [doi: 10.1186/s13223-020-0405-7] [Medline: 32042285]
7. Kayumba K, Nsanzimana S, Binagwaho A, Mugwaneza P, Rusine J, Remera E, et al. TRACnet internet and short message service technology improves time to antiretroviral therapy initiation among HIV-infected infants in Rwanda. Pediatr Infect Dis J 2016 Jul;35(7):767-771 [FREE Full text] [doi: 10.1097/INF.0000000000001153] [Medline: 27031258]
8. Perrier T, DeRenzi B, Anderson R. USSD: the third universal app. In: Proceedings of the 2015 Annual Symposium on Computing for Development. New York, New York, USA: ACM Press; 2015 Presented at: ACM DEV '15: Annual Symposium on Computing for Development; December, 2015; London United Kingdom p. 13-21 URL: http://dl.acm.org/citation.cfm?doid=2830629.2830645 [doi: 10.1145/2830629.2830645]
9. Acharya A, Cunningham K, Manandhar S, Shrestha N, Chen M, Weissman A. Exploring the use of mobile health to improve community-based health and nutrition service utilization in the hills of Nepal: qualitative study. J Med Internet Res 2020 Sep 11;22(9):e17659 [FREE Full text] [doi: 10.2196/17659] [Medline: 32915151]
10. Aker JC, Mbiti IM. Mobile phones and economic development in Africa. J Econ Perspect 2010 Aug 01;24(3):207-232. [doi: 10.1257/jep.24.3.207]
11. The mobile economy sub-Saharan Africa 2019. GSM Association. 2019. URL: https://www.gsma.com/subsaharanafrica/resources/the-mobile-economy-sub-saharan-africa-2019 [accessed 2021-05-25]
12. Nhavoto JA, Grönlund A. Mobile technologies and geographic information systems to improve health care systems: a literature review. JMIR Mhealth Uhealth 2014 May 08;2(2):e21 [FREE Full text] [doi: 10.2196/mhealth.3216] [Medline: 25099368]
13. AIDSinfo. UNAIDS. URL: http://aidsinfo.unaids.org/ [accessed 2021-05-25]
14. Tuberculosis profile: Malawi. World Health Organization. 2020. URL: https://worldhealthorg.shinyapps.io/tb_profiles/?_inputs_&lan=%22EN%22&iso2=%22MW%22&main_tabs=%22est_tab%22 [accessed 2020-08-26]
15. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
16. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. Br Med J 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]
17. Patnaik S, Brunskill E, Thies W. Evaluating the accuracy of data collection on mobile phones: a study of forms, SMS, and voice. In: Proceedings International Conference on Information and Communication Technologies and Development (ICTD). 2009 Presented at: International Conference on Information and Communication Technologies and Development (ICTD); April 17-19, 2009; Doha, Qatar. [doi: 10.1109/ictd.2009.5426700]
18. Airtel integrated report 2019. Airtel Malawi. 2020. URL: https://www.airtel.mw/assets/pdf/investor/Airtel-Malawi-PLC-Annual-Report-2019.pdf [accessed 2021-05-24]
19. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
20. Venter W, Coleman J, Chan VL, Shubber Z, Phatsoane M, Gorgens M, et al. Improving linkage to HIV care through mobile phone apps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e155 [FREE Full text] [doi: 10.2196/mhealth.8376] [Medline: 30021706]
21. Shieshia M, Noel M, Andersson S, Felling B, Alva S, Agarwal S, et al. Strengthening community health supply chain performance through an integrated approach: using mHealth technology and multilevel teams in Malawi. J Glob Health 2014 Dec;4(2):020406 [FREE Full text] [doi: 10.7189/jogh.04.020406] [Medline: 25520796]
22. Barrington J, Wereko-Brobby O, Ward P, Mwafongo W, Kungulwe S. SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malar J 2010 Oct 27;9(1):298 [FREE Full text] [doi: 10.1186/1475-2875-9-298] [Medline: 20979633]
23. Keeton C. Measuring the impact of e-health. Bull World Health Org 2012 May 01;90(5):326-327. [doi: 10.2471/blt.12.020512]
24. Medhi I, Patnaik S, Brunskill E, Gautama SN, Thies W, Toyama K. Designing mobile interfaces for novice and low-literacy users. ACM Trans Comput-Hum Interact 2011 Apr 01;18(1):1-28. [doi: 10.1145/1959022.1959024]
25. Hattie J, Timperley H. The Power of Feedback. Rev Educ Res 2016 Nov 16;77(1):81-112. [doi: 10.3102/003465430298487]

XSL•FO
RenderX

26. Gibson E, Deo S, Jónasson JO, Kachule M, Palamountain K. Redesigning sample transportation in Malawi through improved data sharing and daily route optimization. SSRN J 2020:1-36. [doi: 10.2139/ssrn.3712556]
27. Sutcliffe CG, Thuma PE, van Dijk JH, Sinywimaanzi K, Mweetwa S, Hamahuwa M, et al. Use of mobile phones and text messaging to decrease the turnaround time for early infant HIV diagnosis and notification in rural Zambia: an observational study. BMC Pediatr 2017 Mar 08;17(1):66 [FREE Full text] [doi: 10.1186/s12887-017-0822-z] [Medline: 28270134]
28. Laar A, Bekyieriya E, Isang S, Baguune B. Assessment of mobile health technology for maternal and child health services in rural Upper West Region of Ghana. Public Health 2019 Mar;168:1-8. [doi: 10.1016/j.puhe.2018.11.014] [Medline: 30660898]
29. Lemay NV, Sullivan T, Jumbe B, Perry CP. Reaching remote health workers in Malawi: baseline assessment of a pilot mHealth intervention. J Health Commun 2012;17 Suppl 1:105-117. [doi: 10.1080/10810730.2011.649106] [Medline: 22548604]
30. Déglise C, Suggs LS, Odermatt P. SMS for disease control in developing countries: a systematic review of mobile health applications. J Telemed Telecare 2012 Jul;18(5):273-281. [doi: 10.1258/jtt.2012.110810] [Medline: 22826375]
31. Noordam AC, George A, Sharkey AB, Jafarli A, Bakshi SS, Kim JC. Assessing scale-up of mHealth innovations based on intervention complexity: two case studies of child health programs in Malawi and Zambia. J Health Commun 2015 Jan 12;20(3):343-353. [doi: 10.1080/10810730.2014.965363] [Medline: 25581520]
32. Siedner MJ, Santorino D, Lankowski AJ, Kanyesigye M, Bwana MB, Haberer JE, et al. A combination SMS and transportation reimbursement intervention to improve HIV care following abnormal CD4 test results in rural Uganda: a prospective observational cohort study. BMC Med 2015 Jul 06;13:160 [FREE Full text] [doi: 10.1186/s12916-015-0397-1] [Medline: 26149722]

Abbreviations

EID: early infant diagnosis
mHealth: mobile health
ML: molecular laboratory
MOH: Ministry of Health
R4H: Riders for Health International
ST: sample transportation
TB: tuberculosis
USSD: unstructured supplementary service data
VL: viral load

Edited by R Kukafka; submitted 17.12.20; peer-reviewed by A Lankowski, L Stewart-Isherwood; comments to author 16.01.21; revised version received 13.03.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Killian D, Gibson E, Kachule M, Palamountain K, Bangoh JB, Deo S, Jonasson JO
An Unstructured Supplementary Service Data System for Daily Tracking of Patient Samples and Diagnostic Results in a Diagnostic Network in Malawi: System Development and Field Trial
J Med Internet Res 2021;23(7):e26582
URL: https://www.jmir.org/2021/7/e26582
doi:10.2196/26582
PMID:34255690

©Daniel Killian, Emma Gibson, Mphatso Kachule, Kara Palamountain, Joseph Bitilinyu Bangoh, Sarang Deo, Jonas Oddur Jonasson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

App-Based Feedback for Rehabilitation Exercise Correction in Patients With Knee or Hip Osteoarthritis: Prospective Cohort Study

Johanna Theresia Biebl[1], MD; Marzena Rykala[1], BSc; Maximilian Strobel[2], MSc; Pawandeep Kaur Bollinger[2], PhD; Bernhard Ulm[3], MSc; Eduard Kraft[1], MD; Stephan Huber[2*], MD; Andreas Lorenz[1*], MD

[1]Department of Orthopaedics, Physical Medicine, and Rehabilitation, University Hospital, Ludwig Maximilians University of Munich, Munich, Germany
[2]Kaia Health GmbH, Munich, Germany
[3]Unabhängige statistische Beratung Bernhard Ulm, Munich, Germany
[*]these authors contributed equally

Corresponding Author:
Johanna Theresia Biebl, MD
Department of Orthopaedics, Physical Medicine, and Rehabilitation
University Hospital
Ludwig Maximilians University of Munich
Marchioninistr. 15
Munich, 81377
Germany
Phone: 49 89440074070
Fax: 49 89440077073
Email: johanna.biebl@med.uni-muenchen.de

Abstract

Background: The use of digital therapeutic solutions for rehabilitation of conditions such as osteoarthritis provides scalable access to rehabilitation. Few validated technological solutions exist to ensure supervision of users while they exercise at home. Motion Coach (Kaia Health GmbH) provides audiovisual feedback on exercise execution in real time on conventional smartphones.

Objective: We hypothesized that the interrater agreement between physiotherapists and Motion Coach would be noninferior to physiotherapists' interrater agreement for exercise evaluations in a cohort with osteoarthritis.

Methods: Patients diagnosed with osteoarthritis of the knee or hip were recruited at a university hospital to perform a set of 6 exercises. Agreement between Motion Coach and 2 physiotherapists' corrections for segments of the exercises were compared using Cohen κ and percent agreement.

Results: Participants (n=24) were enrolled and evaluated. There were no significant differences between interrater agreements (Motion Coach app vs physiotherapists: percent agreement 0.828; physiotherapist 1 vs physiotherapist 2: percent agreement 0.833; P<.001). Age (70 years or under, older than 70 years), gender (male, female), or BMI (30 kg/m^2 or under, greater than 30 kg/m^2) subgroup analysis revealed no detectable difference in interrater agreement. There was no detectable difference in levels of interrater agreement between Motion Coach vs physiotherapists and between physiotherapists in any of the 6 exercises.

Conclusions: The results demonstrated that Motion Coach is noninferior to physiotherapist evaluations. Interrater agreement did not differ between 2 physiotherapists or between physiotherapists and the Motion Coach app. This finding was valid for all investigated exercises and subgroups. These results confirm the ability of Motion Coach to detect user form during exercise and provide valid feedback to users with musculoskeletal disorders.

(J Med Internet Res 2021;23(7):e26658) doi:10.2196/26658

KEYWORDS

mHealth; digital health; digital rehabilitation; machine learning; smartphone; osteoarthritis; exercise therapy

XSL•FO
RenderX

Introduction

Musculoskeletal conditions such as osteoarthritis and back pain result in a huge burden for patients and health care systems. Impaired mobility affects both the quality of life of the individual, for example, by increasing social isolation, and the health care system, by raising costs due to factors such as hospitalizations and secondary diseases [1-3]. Osteoarthritis can lead to pain-related fear of movement and an increased probability of further functional impairment [4]. In addition, osteoarthritis is a predictor for developing disabilities that affect activities of daily living, underlining the importance of effective interventions [5].

Current guidelines [6] recommend self-management programs and exercise as first-line therapies for managing osteoarthritis. The prevalence of osteoarthritis is increasing, yet cost and resource constraints limit in-person access to these therapies [7]. Digital therapeutics have emerged as an option to provide access to exercise therapy and multidisciplinary rehabilitation for patients with musculoskeletal pain conditions such as osteoarthritis and back pain [8-10]. Even though a recent survey among health professionals indicated widespread support of use of mobile health technologies in osteoarthritis treatment [11], a primary concern with using digital therapeutics for home-based exercise is the lack of supervision by health care professionals.

Several different digital solutions have been proposed to correct and optimize body pose during exercise execution to improve access to therapeutic exercises [12]. Many mobile health apps for musculoskeletal rehabilitation rely upon video instructions only and provide no means of detecting and correcting pose during exercise [9,13]. These systems, by default, leave users exposed to the risk of incorrectly performing exercise but allow for scalable access without requiring external hardware. To the best of our knowledge, there are no reports on the quality of exercise execution during the use of these systems. Other technologies, such as integrated devices containing inertial sensors, have also been validated to a limited extent, and whether they are suitable for detecting and correcting form during therapeutic exercises has not been evaluated [14,15]. Digital therapeutics that have been validated for this purpose require additional hardware such as a Microsoft Kinect device [16,17].

Motion Coach (Kaia Health GmbH) was recently introduced to address these issues (ie, requiring that equipment be worn on the body or additional hardware) by using only smartphone front camera data and machine learning algorithms to detect the position of body segments during exercise in real time in order to provide personalized feedback.

The aim of this study was to evaluate the ability of Motion Coach to detect and correct form during physiotherapeutic exercises in patients with osteoarthritis. We hypothesized that interrater agreement between physiotherapists and Motion Coach would be noninferior to that between 2 physiotherapists.

Methods

Participants

Participants with a confirmed prior diagnosis of osteoarthritis of the hip or knee were enrolled from the outpatient population of the Department of Orthopedics, Physical Medicine and Rehabilitation, University Hospital, Ludwig Maximilians University of Munich.

Inclusion criteria were (1) diagnosed hip or knee osteoarthritis and (2) age over 18 years. Exclusion criteria were (1) inability to consent (significant cognitive deficits); (2) not fluent in the German language; (3) severe medical or neurological conditions; (4) severe joint contractures that would influence the correct execution of the exercises; (5) previous hip, knee, and ankle arthrodesis; (6) osseous instabilities; or (7) severe osteoporosis.

Ethics and Registration

The study was approved by the Ethics Committee of Ludwig Maximilians University of Munich (20-162) and all participants provided informed consent before study procedures were carried out. The study was registered with the German Study Registry (*Deutsches Register Klinischer Studien*; DRKS00021828) prior to beginning enrollment.

Procedure

To evaluate the correction of osteoarthritis-specific exercises, Motion Coach provides instructions visually through an iPad's screen and acoustically via headphones to the participants. While participants performed exercises using Motion Coach, 2 physiotherapists evaluated whether the exercises were being performed correctly. (Physiotherapists were blinded to the audiovisual feedback of Motion Coach). Furthermore, the physiotherapists evaluated the execution of an exercise set or the performance over the predefined time for static exercises as a whole on a 6-point Likert scale (0=insufficient, 5=excellent execution of movement).

Exercises

For assessment, 6 exercises (Table 1 and Figure 1) that reflected several aspects of therapeutic exercises were chosen from the app to ensure detection by the algorithm was reliable in different circumstances. We included exercises that required a varying range of technical ability; exercises that had different modes of execution (4 dynamic and 2 static), to differentiate between exercises requiring rapid feedback in real time (due to continuous movement) and those that do not; and exercises with different levels of difficulty (low, medium, or high).

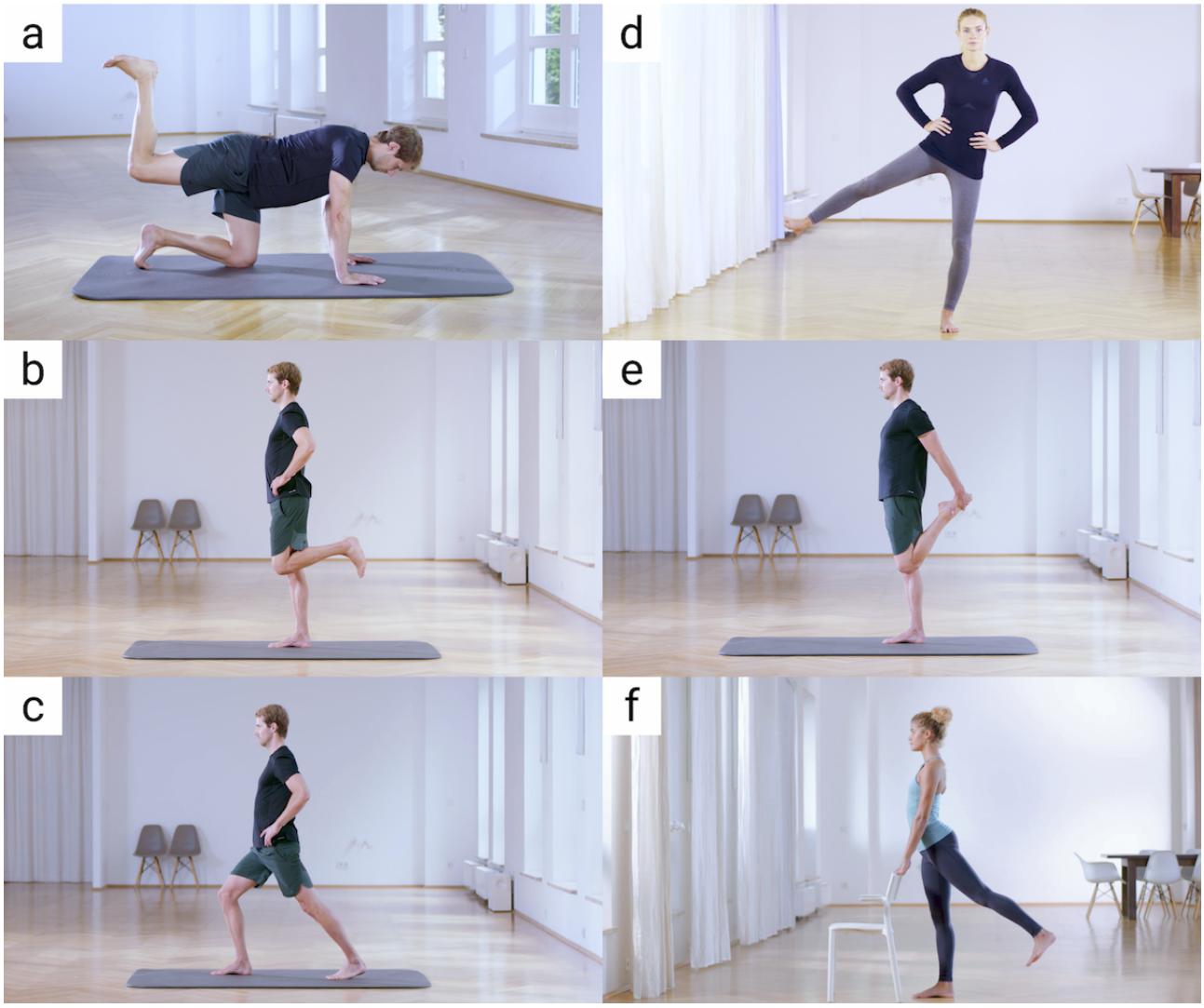
XSL•FO
RenderX

Table 1. Exercises performed by participants. Exercise difficulty was rated by training experts prior to study.

Label[a]	Exercise name	Pose	Execution mode	Exercise difficulty rating
a	Hip extension bent leg	Quadruped	Dynamic	High
b	Knee flexion (leg curl)	Standing	Dynamic	High
c	Strengthening hip extensors	Standing	Dynamic	Medium
d	Strengthen hip abductors	Standing	Dynamic	Medium
e	Strain front of thigh	Standing	Static	Medium
f	Elongation of the hip flexors	Standing	Static	Low

[a]Letters correspond to those in Figure 1.

Figure 1. Exercises performed in this study (a) hip extension bent leg; (b) knee flexion (leg curl); (c) strengthening hip extensors; (d) strengthen hip abductors; (e) strain front of thigh; (f) elongation of the hip flexors.



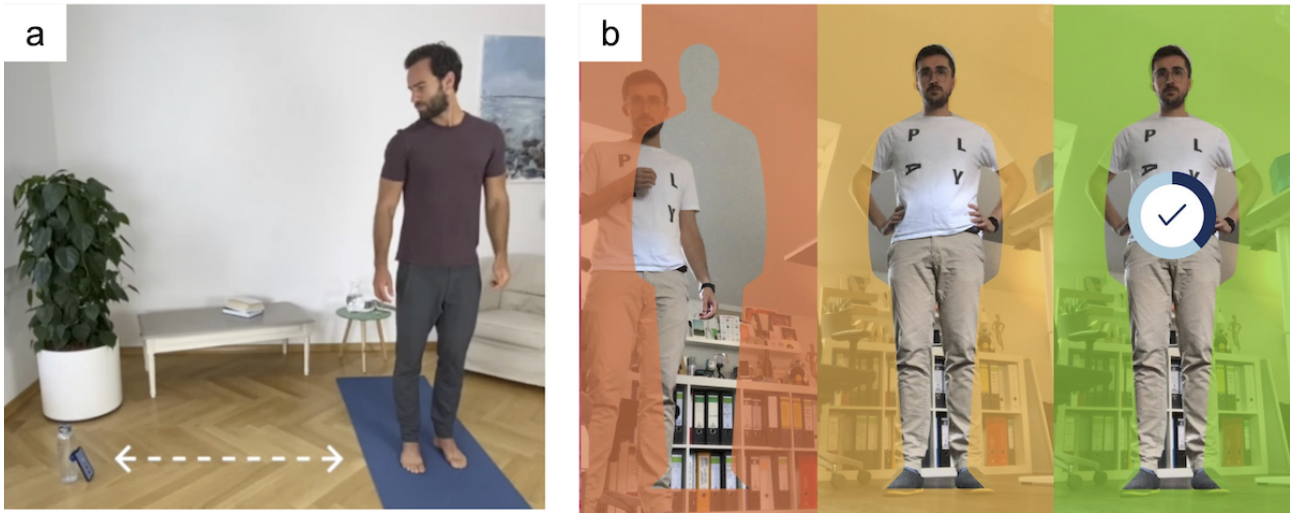
Motion Coach

Overview

In order to give audiovisual feedback on exercise form in real time, Motion Coach uses the camera stream of a user's mobile device and artificial intelligence–based image processing. Users place their device on the ground approximately 2 meters away, tilted slightly so they can be seen in the frame of view of the camera. The app guides the user with interactive setup instructions (Figure 2). A 2-step process is applied to each new image frame as it is captured by the camera.

XSL•FO
RenderX

Figure 2. (a) User stands approximately 2 meters away from their device while the front-facing camera of the device captures user's movements. (b) User is guided as to where to stand by a series of interactive screens.



Step 1: Estimating Pose

First, a Pose Estimation Machine Learning Model is applied to infer the user's pose for each captured image frame in real time (Figure 3). This Pose Estimation Model is a convolutional neural network (typically used for image-based machine learning tasks [18]) with a proprietary architecture that runs entirely on the user's mobile device (therefore, no raw video data leave the user's device). The model was specifically optimized to run on a wide variety of iOS and Android devices, and the model achieves state-of-the-art performance on academic benchmarks such as the MPII Human Pose Data Set Benchmark [19]. Kaia Health trained this model using a proprietary image data set that consisted of data from people with a variety of characteristics (body shape, height, skin color, movement limitations, etc) exercising in front of their mobile device, with a wide variety of exercise movements and environmental conditions such as varying lighting and background to make the model robust. Each image in the data set had been manually labeled according to a taxonomy designed to best capture the human body in physiotherapeutic exercises.

Figure 3. Examples of keypoint poses (white) inferred for various exercises by the Pose Estimation Model.



Step 2: Evaluating Geometric Expert System

For audiovisual feedback, spatiotemporal constraints, which were configured in advance by medical, physiotherapeutic, or sport science–trained Kaia staff, are triggered based on movement; there was no need for reconfiguration on a per-user or per-session basis. While the system was in use, constraints were checked automatically in real time, and feedback was provided if any of the configured constraints were violated. If multiple constraints were violated, the prioritization mechanism selects the feedback based on risk of injury.

Data

Overview

Physiotherapists' evaluations were collected on a rating sheet for each participant. Data from the app were obtained by taking a screenshot of the report of corrections after the exercises had been executed. Baseline data were collected from participants using paper-based surveys or from participants' medical reports if they were available in the system. Data from all sources were entered into a metafile in a spreadsheet (Excel; Microsoft Inc).

XSL•FO
RenderX

Data Collection

Gender, age, diagnosis, location of osteoarthritis, height, weight, and the Western Ontario and McMaster Universities Arthritis Index (WOMAC) score were collected at baseline [20].

Each participant performed 6 exercises with a total of 23 rated segments (a set of repetitions of 10 for each exercise or 30 seconds of stable posing for static exercises). For each segment, each physiotherapist's evaluation and Motion Coach's evaluation (ie, whether correction was required or not) were collected after the participants completed each exercise. Furthermore, the overall form rating by physiotherapists was recorded on a 6-point Likert scale. Data were pooled for the primary analysis.

Study Endpoints

The primary endpoint was overall agreement between physiotherapists' and Motion Coach' evaluations during exercise execution. For each segment, there was a dichotomous outcome (correction recommended or not).

Sample Size

We calculated the sample size required for a noninferiority trial with dichotomous outcome (ie, agreement or disagreement, either between app and physiotherapists' ratings or between the 2 physiotherapists). We used pilot data (app–physiotherapists mean ratio 0.83; physiotherapist 1–physiotherapist 2 mean ratio 0.845) from the first 16 participants of the study. We determined that 552 exercise segments would be required; therefore, given an assumption of 23 segments per participant, the number of required participants was 24 (noninferiority margin 0.05; α=5%; β=90%). A noninferiority margin of 0.05 was recently used in a comparable study [16] for evaluation of exercise correction with a digital tool.

Statistical Analysis

Continuous data (age, weight, height, and BMI) are described using means and standard deviations; discrete data (gender, location of osteoarthritis, WOMAC score) are described using absolute and relative numbers. Motion Coach–physiotherapist 1, Motion Coach–physiotherapist 2, Motion Coach–both physiotherapists, and physiotherapist 1–physiotherapist 2 interrater reliabilities (Cohen κ and percent agreement) were compared using z scores (α=5%). To assess whether demographic variables had any significant effect on the interrater agreement between Motion Coach and physiotherapists, subgroups for age (70 years or under, older than 70 years), gender (male, female), and BMI (30 kg/m^2 or under, greater than 30 kg/m^2) were formed and compared. We also assessed interrater agreement by exercise. Interrater agreement was categorized according to Cohen κ values as suggested by Landis and Koch [21]: κ < 0.00, poor agreement; κ=0.00-0.20, slight agreement; κ=0.21-0.40, fair agreement; κ=0.41-0.60, moderate agreement; κ=0.61-0.80, substantial agreement; κ=0.81-1.00, almost perfect agreement. All analyses were conducted with R software (version 4.0.2; R Foundation for Statistical Computing).

Results

Participants

The study population's mean age was 67.6 (SD 8.98 years), and 20 out of the 24 participants (83%) were female. Participants (Table 2) had osteoarthritis of the knee (15/24, 62.5%), hip (6/24, 25%), or both knee and hip (3/24, 12.5%).

The mean global WOMAC score was 64.9 (SD 43.3) with mean domain scores of 15.8 (SD 10.7) for Pain, 7.3 (SD 4.8) for Stiffness, and 41.9 (SD 30.5) for Physical Function.

XSL•FO
RenderX

Table 2. Study population characteristics.

Characteristic	Value (n=24)
Gender, n (%)	
Male	4 (17)
Female	20 (83)
Age (years)	
mean (SD)	67.6 (9.0)
n (%)	
≤70	12 (50)
>70 years	12 (50
Weight (kg), mean (SD)	69.5 (16.7)
Height (m), mean (SD)	1.7 (0.1)
BMI (kg/m^2)	
mean (SD)	24.9 (4.6)
n (%)	
≤30 kg/m^2	20
>30 kg/m^2	4
Location of osteoarthritis, n (%)	
Hip	6 (25)
Knee	15 (63)
Both hip and knee	3 (13)
WOMAC[a], n (%)	
Total score	65 (43)
Pain	16 (11)
Stiffness	7 (5)
Physical function	42 (31)

[a]WOMAC: Western Ontario and McMaster Universities Arthritis Index.

Primary Analysis

Mean agreement between the app and physiotherapists (percent agreement 0.828) was not inferior (margin 0.05; *P*<.001) to that between physiotherapist 1 and physiotherapist 2 (percent agreement 0.833).

Comparison of Interrater Reliability

Interrater reliability for the evaluations (Table 3) demonstrated moderate to substantial agreement between physiotherapist 1 and physiotherapist 2 (Cohen κ=0.607, 95% CI 0.535-0.679; percent agreement 0.833, 95% CI 0.800-0.864), Motion Coach and physiotherapist 1 (Cohen κ=0.551, 95% CI 0.474-0.628; percent agreement 0.815, 95% CI 0.780-0.847); Motion Coach and physiotherapist 2 (Cohen κ=0.626, 95% CI 0.474-0.697; percent agreement 0.841, 95% CI 0.807-0.870); and Motion Coach and when there was agreement between both physiotherapists (Cohen κ=0.726, 95% CI 0.654-0.798; percent agreement 0.893, 95% CI 0.862-0.920) (Figure 3). There was no detectable difference between either Motion Coach–physiotherapist 1 interrater reliability and physiotherapist 1–physiotherapist 2 interrater reliability (Cohen κ: *P*=.309; percent agreement: *P*=.46) or Motion Coach–physiotherapist interrater reliability and physiotherapist 1–physiotherapist 2 interrater reliability (Cohen κ: *P*=.71; percent agreement: *P*=.74; Figure 4).

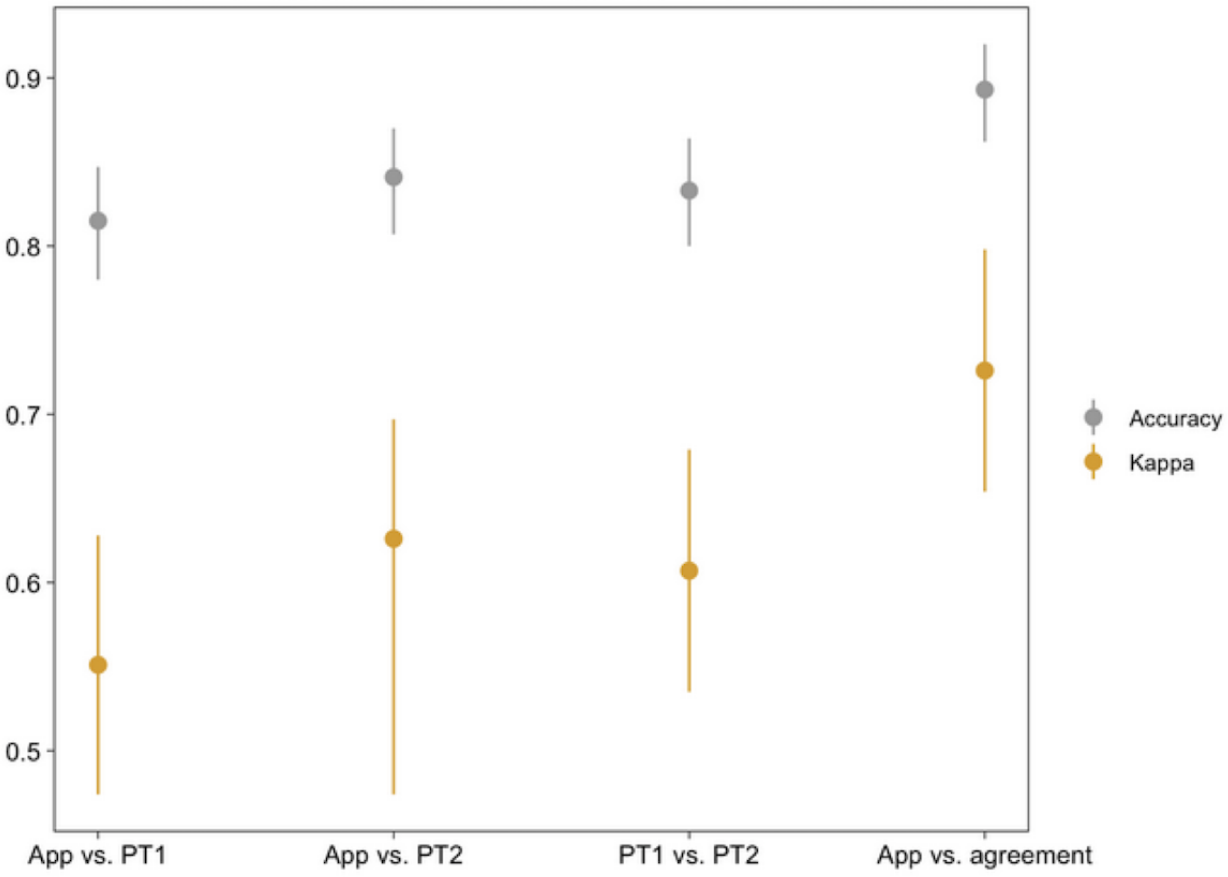
XSL•FO
RenderX

Table 3. All interpretations of correct versus incorrect exercise execution.

Assessment	All	Individual				Both		
		Physiotherapist 1		Physiotherapist 2		Agreement		Disagreement
		Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	
All, n	552	394	158	374	178	338	122	92
Physiotherapist 1, n (%)								
Correct	394 (71.4)	N/A[a]	N/A	338 (90.4)	56 (31.5)	N/A	N/A	N/A
Incorrect	158 (28.6)	N/A	N/A	36 (9.6)	122 (68.5)	N/A	N/A	N/A
App, n (%)								
Correct	390 (70.7)	341 (86.5)	49 (31.0)	338 (90.4)	52 (29.2)	314 (92.9)	25 (20.5)	51 (55.4)
Incorrect	162 (29.3)	53 (13.5)	109 (69.0)	36 (9.6)	126 (70.8)	24 (7.1)	97 (79.5)	41 (44.6)

[a]N/A: not applicable.

Figure 4. Interrater reliability (percent agreement and Cohen κ, with upper and lower 95% confidence intervals). PT: physiotherapist.



Subgroup Analysis

No differences were found between app–physiotherapist interrater reliabilities and physiotherapist 1–physiotherapist 2 interrater reliability in any of the subgroups (Table 4 and Figure 5).

XSL•FO
RenderX

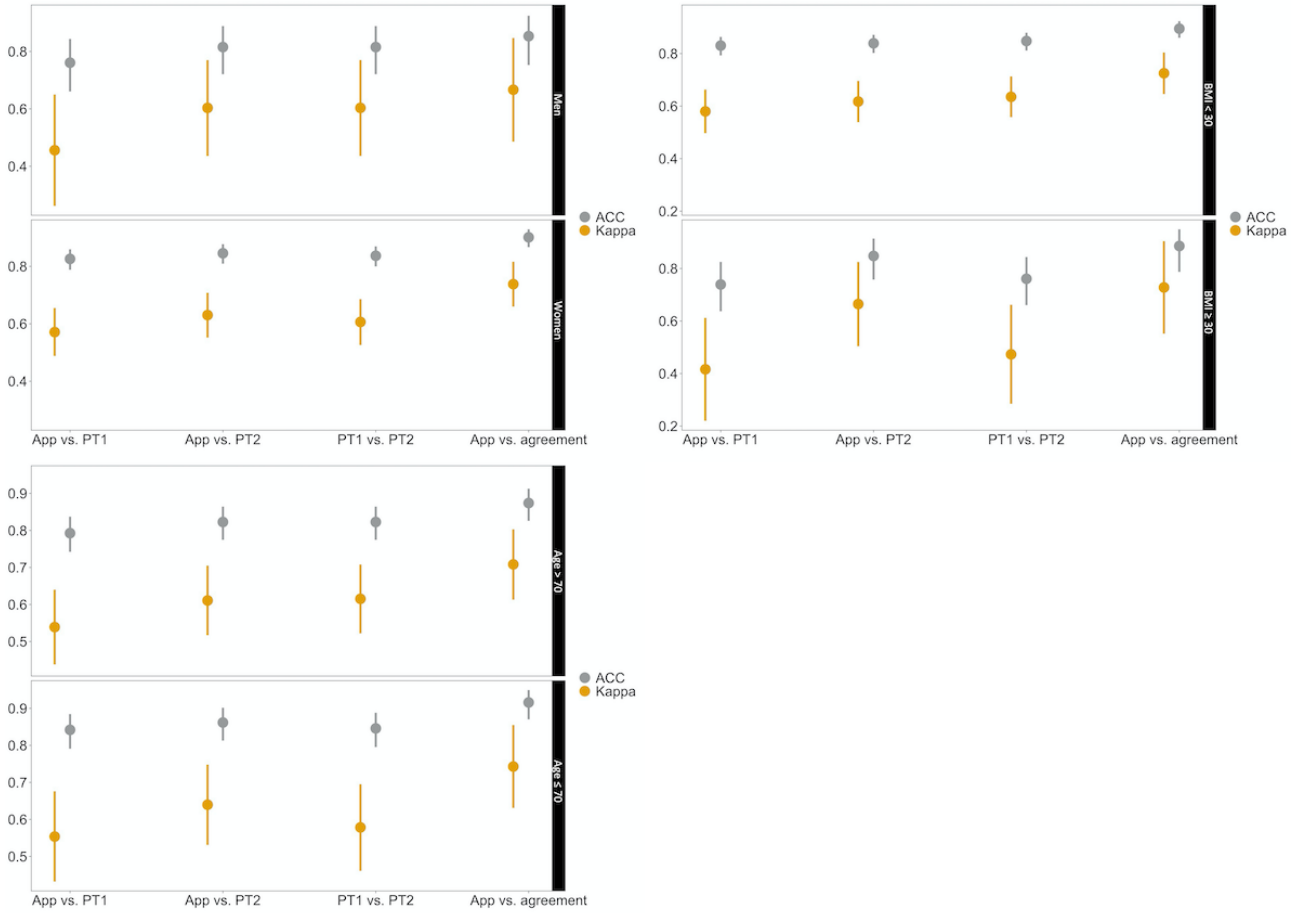
Table 4. Interrater agreement for age, gender, and BMI subgroups.

Raters	Cohen κ		Percent agreement	
	Mean (95% CI)	*P* value[a]	Mean (95% CI)	*P* value[a]
All				
Physiotherapist 1 vs physiotherapist 2	0.607 (0.535-0.679)		0.833 (0.800-0.863)	
App vs physiotherapist 1	0.551 (0.474-0.628)	.31	0.815 (0.780-0.847)	.46
App vs physiotherapist 2	0.626 (0.556-0.697)	.71	0.841 (0.807-0.870)	.74
App vs agreement	0.726 (0.654-0.798)		0.893 (0.862-0.920)	
Gender				
Male				
Physiotherapist 1 vs physiotherapist 2	0.603 (0.436-0.770)		0.815 (0.721-0.889)	
App vs physiotherapist 1	0.456 (0.262-0.650)	.26	0.761 (0.661-0.844)	.34
App vs physiotherapist 2	0.603 (0.436-0.770)	>.999	0.815 (0.721-0.889)	>.999
App vs agreement	0.667 (0.486-0.847)		0.853 (0.753-0.924)	
Female				
Physiotherapist 1 vs physiotherapist 2	0.606 (0.526-0.686)		0.837 (0.800-0.870)	
App vs physiotherapist 1	0.571 (0.488-0.655)	.55	0.826 (0.788-0.860)	.65
App vs physiotherapist 2	0.630 (0.552-0.708)	.67	0.846 (0.809-0.877)	.71
App vs agreement	0.738 (0.660-0.816)		0.901 (0.867-0.929)	
BMI				
<30 kg/m^2				
Physiotherapist 1 vs physiotherapist 2	0.635 (0.558-0.713)		0.848 (0.812-0.879)	
App vs physiotherapist 1	0.580 (0.497-0.663)	.34	0.830 (0.793-0.864)	.46
App vs physiotherapist 2	0.617 (0.539-0.696)	.75	0.839 (0.802-0.872)	.71
App vs agreement	0.725 (0.646-0.804)		0.895 (0.860-0.923)	
≥30 kg/m^2				
Physiotherapist 1 vs physiotherapist 2	0.473 (0.285-0.662)		0.761 (0.661-0.844)	
App vs physiotherapist 1	0.416 (0.220-0.612)	.68	0.739 (0.637-0.825)	.72
App vs physiotherapist 2	0.665 (0.504-0.825)	.13	0.848 (0.758-0.914)	.11
App vs agreement	0.728 (0.552-0.904)		0.886 (0.787-0.949)	
Age				
≤70 years				
Physiotherapist 1 vs physiotherapist 2	0.578 (0.461-0.695)		0.846 (0.795-0.888)	
App vs physiotherapist 1	0.554 (0.432-0.676)	.77	0.842 (0.791-0.885)	.90
App vs physiotherapist 2	0.640 (0.531-0.748)	.45	0.862 (0.813-0.902)	.59
App vs agreement	0.743 (0.631-0.855)		0.916 (0.870-0.949)	
>70 years				
Physiotherapist 1 vs physiotherapist 2	0.615 (0.522-0.708)		0.823 (0.775-0.864)	
App vs physiotherapist 1	0.539 (0.438-0.640)	.28	0.793 (0.742-0.837)	.33
App vs physiotherapist 2	0.611 (0.517-0.705)	.95	0.823 (0.775-0.864)	>.999
App vs agreement	0.708 (0.613-0.803)		0.874 (0.826-0.913)	

[a]Comparison of subrow with *Physiotherapist 1 vs physiotherapist 2*.

XSL•FO
RenderX

Figure 5. Interrater reliability (percent agreement and Cohen κ, with upper and lower 95% confidence intervals) for (a) gender, (b) BMI, and (c) age subanalyses. PT: physiotherapist.



Interrater Agreement in Different Exercises

The analysis showed no detectable difference in the rates of interrater agreement in any of the exercises (Table 5 and Table 6).

Table 5. Mean rating of exercise form by the physiotherapists, using a 6-point Likert scale, and interrater agreement comparisons between app–physiotherapist and physiotherapist 1–physiotherapist 2 percent agreement values for each exercise.

Exercise[a]	Rating, mean (SD)	Percent agreement				*P* value	
		App–physiotherapist 1	App–physiotherapist 2	Physiotherapist 1–physiotherapist 2	App–agreement	Comparison 1[b]	Comparison 2[c]
a	2.8 (1.1)	0.804 (0.735-0.861)	0.845 (0.782-0.896)	0.851 (0.788-0.896)	0.881 (0.816-0.929)	.32	.23
b	3.4 (1.3)	0.792 (0.680-0.878)	0.806 (0.695-0.889)	0.875 (0.776-0.889)	0.841 (0.727-0.921)	.86	.13
c	4.3 (1.4)	0.778 (0.664-0.867)	0.750 (0.634-0.845)	0.750 (0.634-0.845)	0.852 (0.729-0.934)	.67	.67
d	4.5 (1.1)	0.889 (0.793-0.951)	0.889 (0.793-0.951)	0.889 (0.793-0.951)	0.938 (0.840-0.983)	>.999	>.999
e	4.5 (1.2)	0.833 (0.727-0.911)	0.861 (0.759-0.931)	0.778 (0.664-0.931)	0.946 (0.851-0.989)	.48	.36
f	4.8 (1.0)	0.812 (0.720-0.885)	0.875 (0.792-0.934)	0.833 (0.744-0.934)	0.912 (0.828-0.964)	.12	.68

[a]Letters correspond to those in Figure 1.

[b]App–physiotherapist 1 vs physiotherapist 1–physiotherapist 2.

[c]App–physiotherapist 2 vs physiotherapist 1–physiotherapist 2.

XSL•FO
RenderX

Table 6. Mean rating of exercise form by physiotherapists, using a 6-point Likert scale, and comparisons between app–physiotherapist and physiotherapist 1–physiotherapist 2 Cohen κ values for each exercise.

Exercise[a]	Rating, mean (SD)	Cohen κ				*P* value	
		App–physiotherapist 1	App–physiotherapist 2	Physiotherapist 1–physiotherapist 2	App–agreement	Comparison 1[b]	Comparison 2[c]
a	2.8 (1.1)	0.534 (0.396-0.673)	0.655 (0.396-0.673)	0.656 (0.396-0.673)	.707 (0.396-0.673)	.20	.20
b	3.4 (1.3)	0.579 (0.391-0.766)	0.605 (0.391-0.766)	0.749 (0.391-0.766)	.679 (0.391-0.766)	.85	.17
c	4.3 (1.4)	0.464 (0.242-0.687)	0.397 (0.242-0.687)	0.425 (0.242-0.687)	.596 (0.242-0.687)	.68	.81
d	4.5 (1.1)	0.680 (0.476-0.884)	0.695 (0.476-0.884)	0.714 (0.476-0.884)	.817 (0.476-0.884)	.92	.81
e	4.5 (1.2)	0.597 (0.393-0.801)	0.681 (0.393-0.801)	0.478 (0.393-0.801)	.867 (0.393-0.801)	.55	.44
f	4.8 (1.0)	0.435 (0.218-0.651)	0.635 (0.218-0.651)	0.453 (0.218-0.651)	.616 (0.218-0.651)	.17	.91

[a]Letters correspond to those in Figure 1.

[b]App–physiotherapist 1 vs physiotherapist 1–physiotherapist 2.

[c]App–physiotherapist 2 vs physiotherapist 1–physiotherapist 2.

Discussion

The purpose of this study was to compare interrater agreement of osteoarthritis knee and hip exercise assessments between Motion Coach (a novel digital tool) and trained physiotherapists; we hypothesized that assessment agreement for the Motion Coach app would not be inferior to that of physiotherapists. Our data support the hypothesis that Motion Coach is noninferior to physiotherapists in assessing whether exercise poses required correction. There was no difference between the interrater agreement of Motion Coach and physiotherapists and that among physiotherapists. This finding was also true in analyses of subgroups that consisted of men, women, participants 70 years or older, participants below 70 years, participants with BMI greater than 30 kg/m^2, and participants with BMI less than 30 kg/m^2 and in analyses by exercise. To the best of our knowledge, this is the first report comparing a digital software–based exercise feedback tool with conventional smartphone technology and physiotherapeutic exercise feedback for musculoskeletal conditions.

Previous studies [16,17] have used 3D sensors such as the Microsoft Kinect system to assess pose during exercise and give feedback to users if correction was needed. However, 3D-sensor systems are expensive and require extensive external hardware and a stationary television set, and thus have limited scalability in providing access to digital rehabilitation. Komatireddy [16] found no detectable difference in agreement between a software solution for Microsoft Kinect and a panel of physiotherapists for repetition count and the number of acceptable exercises. Wochartz et al [17] evaluated agreement with regard to joint angles and positions of the lower limb between a Microsoft Kinect based-system and a 3D camera-based motion system but did not evaluate its capacity to trigger corrections during therapeutic exercises; they concluded that the validity of the Kinect system to detect pose without postprocessing was restricted.

Other digital rehabilitation tools for musculoskeletal pain use external inertial sensors attached to specific limbs or joints to detect exercise poses [22-24]. By nature, these systems are limited to detecting the poses of joints or body areas only where they are placed, and users must typically attach the hardware to their bodies themselves. Studies [14,15] have shown that these systems are generally capable of detecting exercise poses; however, these systems have not been systematically evaluated for their ability to provide feedback on pose during exercise execution.

Built-in smartphone inertia sensors are a viable option to deliver pose correction in rehabilitation without requiring specialized equipment or installations. Spina et al evaluated real-time smartphone motion sensor data processing as an option to assess pose in physical exercises by people with chronic obstructive pulmonary disease [25]. The system was able to provide feedback on pose and exercise feedback similar to the feedback of a trained therapist. The system required a holster to hold the smartphone and that was repositioned on the body depending on the exercise performed. While previous reports have addressed the general feasibility of exercise-related feedback using 2D RGB camera streams, the percent agreement of those systems without postprocessing limited their use [26,27]. In contrast, Motion Coach relies upon 2D camera stream postprocessing of using machine learning algorithms for valid real-time feedback for exercise correction.

To the best of our knowledge, this study is the first to evaluate the potential of a technology (Motion Coach) to trigger suitable corrections of therapeutic exercises in musculoskeletal pain rehabilitation, with the findings suggesting that Motion Coach technology triggers valid corrections as compared to trained physiotherapists. Motion Coach is a software only solution operating on off-the-shelf smartphones, without any need for additional hardware, which makes this digital therapeutic solution accessible to a broad patient population.

The interrater reliability of trained physiotherapists assessments of pose during lower extremity exercises for the has been investigated: Chmielewski et al [28] investigated interrater agreement during 2 exercises performed by healthy volunteers for the lower extremity with 2 distinct methods (overall rating

XSL•FO
RenderX

and investigation of deviation from the neutral plane during exercise) in a panel of 3 physiotherapists and found agreement better than chance but no high levels of agreement between physiotherapists. Whatman et al [29] investigated interrater agreement for lower extremity exercises in a panel of physiotherapists (segment-specific and overall agreement) with ordinal and dichotomous outcomes; interrater agreement was generally fair to good and increased with experience of the rater. The interrater agreement observed in our study, among the physiotherapists and also between the physiotherapists and Motion Coach, was high compared to those in previous studies [28,29]. This finding can be explained by the high level of experience of the physiotherapists and training of the physiotherapists on evaluation criteria prior to patient enrollment. Compared to other approaches requiring specialized hardware, the degree of agreement between both physiotherapists and Motion Coach remains high; a similar study [30] using data from the Kinect version 2 Skeleton Tracking system to assess rehabilitation exercises in 19 people with musculoskeletal and neurological limitations showed a limited correlation (r=0.60, P<.01 for the clinical subgroup) between expert's clinical judgement and the results of various models based on sensor data.

The study had several limitations. First, the pool of raters was small with n=2, and a third rater was not used (in cases of disagreement between the 2 raters). In addition, the sample was heterogeneous in terms of gender distribution and localization of osteoarthritis, limiting the generalizability of the results. Other limitations arise from the fact that the assessment of pose during therapeutic exercise execution is not standardized, and thus, in this study as in comparable previous studies [28,29], no well-established standard measurement could be used to quantify exercise execution. Furthermore, dichotomous assessment of acceptable exercise is only one of several measures used in prior studies to assess form during exercise. Future studies evaluating Motion Coach will need to use more diverse outcome measures of form during exercise, for example calculations with a musculoskeletal human model.

The interrater agreement for suggesting corrections during therapeutic exercises between both physiotherapists and Motion Coach was moderate to substantial and did not differ between physiotherapists themselves and physiotherapists and Motion Coach. This finding was valid for all investigated exercises and subgroup analysis. These findings validate the ability of Motion Coach to detect form during exercise and provide audiovisual feedback to users with preexisting musculoskeletal conditions.

Acknowledgments

The study was part of a project (TELE-CORRECT, grant number MED-1804-0003) funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy. The authors would like to thank Luca Sonntag at Kaia Health for providing and editing images used in the manuscript.

Conflicts of Interest

SH, PKB, and MS are employees of Kaia Health and receive salary and stock options.

References

1. GBD 2016 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
2. Dawson J, Linsell L, Zondervan K, Rose P, Carr A, Randall T, et al. Impact of persistent hip or knee pain on overall health status in elderly people: a longitudinal population study. Arthritis Rheum 2005 Jun 15;53(3):368-374 [FREE Full text] [doi: 10.1002/art.21180] [Medline: 15934104]
3. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
4. Heuts PH, Vlaeyen JW, Roelofs J, de Bie RA, Aretz K, van Weel C, et al. Pain-related fear and daily functioning in patients with osteoarthritis. Pain 2004 Jul;110(1-2):228-235. [doi: 10.1016/j.pain.2004.03.035] [Medline: 15275772]
5. Song J, Chang RW, Dunlop DD. Population impact of arthritis on disability in older adults. Arthritis Rheum 2006 Apr 15;55(2):248-255 [FREE Full text] [doi: 10.1002/art.21842] [Medline: 16583415]
6. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, American College of Rheumatology. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012 Apr 27;64(4):465-474 [FREE Full text] [doi: 10.1002/acr.21596] [Medline: 22563589]
7. Postler A, Ramos AL, Goronzy J, Günther KP, Lange T, Schmitt J, et al. Prevalence and treatment of hip and knee osteoarthritis in people aged 60 years or older in Germany: an analysis based on health insurance claims data. Clin Interv Aging 2018;13:2339-2349 [FREE Full text] [doi: 10.2147/CIA.S174741] [Medline: 30532524]
8. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]

XSL•FO
RenderX

9. Dahlberg LE, Dell'Isola A, Lohmander LS, Nero H. Improving osteoarthritis care by digital means - effects of a digital self-management program after 24- or 48-weeks of treatment. PLoS One 2020 Mar 4;15(3):e0229783 [FREE Full text] [doi: 10.1371/journal.pone.0229783] [Medline: 32130276]
10. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
11. Biebl J, Huber S, Rykala M, Kraft E, Lorenz A. Attitudes and expectations of health care professionals toward app-based therapy in patients with osteoarthritis of the hip or knee: questionnaire study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e21704 [FREE Full text] [doi: 10.2196/21704] [Medline: 33112255]
12. Liao Y, Vakanski A, Xian M, Paul D, Baker R. A review of computational approaches for evaluation of rehabilitation exercises. Comput Biol Med 2020 Apr;119:103687 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103687] [Medline: 32339122]
13. Irvine AB, Russell H, Manocchia M, Mino DE, Cox GT, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
14. Bauer C, Rast F, Ernst M, Kool J, Oetiker S, Rissanen S, et al. Concurrent validity and reliability of a novel wireless inertial measurement system to assess trunk movement. J Electromyogr Kinesiol 2015 Oct;25(5):782-790. [doi: 10.1016/j.jelekin.2015.06.001] [Medline: 26126796]
15. Bauer CM, Heimgartner M, Rast FM, Ernst MJ, Oetiker S, Kool J. Reliability of lumbar movement dysfunction tests for chronic low back pain patients. Man Ther 2016 Aug;24:81-84. [doi: 10.1016/j.math.2016.02.013] [Medline: 26980560]
16. Komatireddy R, Chokshi A, Basnett J, Casale M, Goble D, Shubert T. Quality and quantity of rehabilitation exercises delivered by a 3-d motion controlled camera: a pilot study. Int J Phys Med Rehabil 2014 Aug;2(4):1 [FREE Full text] [doi: 10.4172/2329-9096.1000214] [Medline: 26824053]
17. Wochatz M, Tilgner N, Mueller S, Rabe S, Eichler S, John M, et al. Reliability and validity of the Kinect V2 for the assessment of lower extremity rehabilitation exercises. Gait Posture 2019 May;70:330-335. [doi: 10.1016/j.gaitpost.2019.03.020] [Medline: 30947108]
18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
19. Parekh P, Patel A. Deep learning-based 2D and 3D human pose estimation: a survey. In: Singh PK, Wierzchoń ST, Tanwar S, Ganzha M, Rodrigues JJPC, editors. Lecture Notes in Networks and Systems, vol 203. Singapore: Springer; 2021.
20. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988 Dec;15(12):1833-1840. [Medline: 3068365]
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
22. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156. [doi: 10.2196/jmir.9667] [Medline: 29695370]
23. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
24. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111. [doi: 10.2196/13111] [Medline: 30816849]
25. Spina G, Huang G, Vaes A, Spruit M, Amft O. COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback. In: Proceedings of the 2013 ACN International Joint Conference on Pervasive and Ubiquitous Computing. 2013 Presented at: 2013 ACN International Joint Conference on Pervasive and Ubiquitous Computing; September 8-12; Zurich, Switzerland p. 597-606. [doi: 10.1145/2493432.2493454]
26. Ar I, Akgul YS. A computerized recognition system for the home-based physiotherapy exercises using an rgbd camera. IEEE Trans Neural Syst Rehabil Eng 2014 Nov;22(6):1160-1171. [doi: 10.1109/tnsre.2014.2326254] [Medline: 24860037]
27. Sucar L, Luis R, Leder R, Hernandez J, Sanchez I. Gesture therapy: a vision-based system for upper extremity stroke rehabilitation. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:3690-3693. [doi: 10.1109/IEMBS.2010.5627458] [Medline: 21096856]
28. Chmielewski TL, Hodges MJ, Horodyski M, Bishop MD, Conrad BP, Tillman SM. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. J Orthop Sports Phys Ther 2007 Mar;37(3):122-129. [doi: 10.2519/jospt.2007.2457] [Medline: 17416127]
29. Whatman C, Hing W, Hume P. Physiotherapist agreement when visually rating movement quality during lower extremity functional screening tests. Phys Ther Sport 2012 May;13(2):87-96. [doi: 10.1016/j.ptsp.2011.07.001] [Medline: 22498149]

XSL•FO
RenderX

30. Capecci M, Ceravolo MG, Ferracuti F, Iarlori S, Kyrki V, Monteriù A, et al. A hidden semi-Markov model based approach for rehabilitation exercise assessment. J Biomed Inform 2018 Feb;78:1-11. [doi: 10.1016/j.jbi.2017.12.012] [Medline: 29277330]

Abbreviations

WOMAC: Western Ontario and McMaster Universities Arthritis Index

Edited by R Kukafka; submitted 20.12.20; peer-reviewed by AV Das, S Kriventsov; comments to author 19.01.21; revised version received 03.02.21; accepted 19.04.21; published 13.07.21.

Please cite as:
Biebl JT, Rykala M, Strobel M, Kaur Bollinger P, Ulm B, Kraft E, Huber S, Lorenz A
App-Based Feedback for Rehabilitation Exercise Correction in Patients With Knee or Hip Osteoarthritis: Prospective Cohort Study
J Med Internet Res 2021;23(7):e26658
URL: https://www.jmir.org/2021/7/e26658
doi:10.2196/26658
PMID:34255677

©Johanna Theresia Biebl, Marzena Rykala, Maximilian Strobel, Pawandeep Kaur Bollinger, Bernhard Ulm, Eduard Kraft, Stephan Huber, Andreas Lorenz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptability of an mHealth App That Provides Harm Reduction Services Among People Who Inject Drugs: Survey Study

Tyler Shelby[1,2], MPhil; Xin Zhou[3], PhD; Douglas Barber[1], BA; Frederick Altice[2,3,4], MD

[1]Yale University School of Medicine, New Haven, CT, United States

[2]Department of Epidemiology of Microbial Diseases, Yale University School of Public Health, New Haven, CT, United States

[3]Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT, United States

[4]Centre of Excellence on Research on AIDS, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Tyler Shelby, MPhil
Department of Epidemiology of Microbial Diseases
Yale University School of Public Health
60 College Street
New Haven, CT, 06510
United States
Phone: 1 6202284003
Email: tyler.shelby@yale.edu

Abstract

Background: Harm reduction services reduce the negative consequences of drug injection and are often embedded within syringe service programs (SSPs). However, people who inject drugs (PWID) suboptimally engage with such services because of stigma, fear, transportation restrictions, and limited hours of operation. Mobile health (mHealth) apps may provide an opportunity to overcome these barriers and extend the reach of SSPs beyond that of the traditional brick-and-mortar models.

Objective: This study aims to assess the prevalence of smartphone ownership, the level of comfort in providing the personal information required to use mHealth apps, and interest in using an mHealth app to access harm reduction services among PWID to guide the development of an app.

Methods: We administered a survey to 115 PWID who were enrolled via respondent-driven sampling from July 2018 to July 2019. We examined the extent to which PWID had access to smartphones; were comfortable in providing personal information such as name, email, and address; and expressed interest in various app-based services. We measured participant characteristics (demographics, health status, and behaviors) and used binary logistic and Poisson regressions to identify independent correlates of mHealth-related variables. The primary regression outcomes included summary scores for *access*, *comfort*, and *interest*. The secondary outcomes included binary survey responses for individual comfort or interest components.

Results: Most participants were White (74/105, 70.5%), male (78/115, 67.8%), and middle-aged (mean=41.7 years), and 67.9% (74/109) owned a smartphone. Participants reported high levels of comfort in providing personal information to use an mHealth app, including name (96/109, 88.1%), phone number (92/109, 84.4%), email (85/109, 77.9%), physical address (85/109, 77.9%), and linkage to medical records (72/109, 66.1%). Participants also reported strong interest in app-based services, including medication or sterile syringe delivery (100/110, 90.9%), lab or appointment scheduling (90/110, 81.8%), medication reminders (77/110, 70%), educational material (65/110, 59.1%), and group communication forums (64/110, 58.2%). Most participants were comfortable with the idea of home delivery of syringes (93/109, 85.3%). Homeless participants had lower access to smartphones (adjusted odds ratio [AOR] 0.15, 95% CI 0.05-0.46; *P*=.001), but no other participant characteristics were associated with primary outcomes. Among secondary outcomes, recent SSP use was positively associated with comfort with the home delivery of syringes (AOR 3.29, 95% CI 1.04-10.3 *P*=.04), and being older than 50 years was associated with an increased interest in educational materials (AOR 4.64, 95% CI 1.31-16.5; *P*=.02) and group communication forums (AOR 3.69, 95% CI 1.10-12.4; *P*=.04).

Conclusions: Our findings suggest that aside from those experiencing homelessness or unstable housing, PWID broadly have access to smartphones, are comfortable with sharing personal information, and express interest in a wide array of services within an app. Given the suboptimal access to and use of SSPs among PWID, an mHealth app has a high potential to address the harm reduction needs of this vulnerable population.

(J Med Internet Res 2021;23(7):e25428) doi:10.2196/25428

XSL•FO
RenderX

KEYWORDS

people who inject drugs; mHealth; patient preferences; syringe services programs; service access; mobile phone

Introduction

Background

The current opioid epidemic has resulted in many adverse outcomes, including increases in hospitalizations attributed to opioid injection and increases in infections of hepatitis C virus (HCV) and HIV, endocarditis, soft tissue infections, and overdose death [1-4]. Reductions in HIV incidence have stalled as a consequence of the volatile opioid epidemic, and numerous outbreaks of HIV and HCV have been reported in both urban and rural settings [4,5]. Although the increasing rates of infection among people who inject drugs (PWID) are staggering, approximately 130 individuals die from overdose each day in the United States [6]. This ongoing crisis calls for further implementation and development of novel methods to reduce risk among PWID and increase the delivery of harm reduction services.

To reduce the incidence of such negative health outcomes, syringe service programs (SSPs) offer services including syringe exchange; the provision of clean injection supplies, fentanyl test kits, and bandages; naloxone (Narcan) training and kit provision; and linkage to internal or external programs for housing support, mental health counseling, primary care, and addiction treatment. Through such a wide range of services, SSPs have demonstrated efficacy in reducing rates of HIV transmission [7] and syringe sharing [8] and increasing HIV treatment and prevention cascades [9-11]. However, SSPs' uptake and coverage are far from meeting recommended targets and provide consistent services to only a quarter of PWID [12]. Reasons for low uptake include (1) long geographical distance to SSPs [13-15], especially in rural areas where opioid use is rising substantially [15]; (2) limited SSP hours of operation [16]; (3) lack of public transportation [17]; (4) the limited power of SSP workforces [18]; and (5) perceived stigma among PWID and fear of arrest and police interference [8,19-22]. These reasons may largely be attributed to the operational style of current SSPs, that is, services and supplies are either provided at central SSP locations or delivered through mobile vans at selected sites during scheduled hours [23]. Although home delivery and contact-free SSPs have been suggested as alternatives to overcome these barriers, they remain underexplored [24,25].

With regard to improving the accessibility and coverage of harm reduction services, a mobile health (mHealth) app has great potential to provide PWID with better access to health care and harm reduction services while also protecting their privacy and offering them a better sense of control of their environment. PWID increasingly have access to smartphones [26], and mHealth interventions have demonstrated significant potential to positively impact a variety of health conditions, including obesity prevention [27], physical activity and healthy eating promotion [28,29], cerebral stroke detection and management [30], and diabetes management [31,32], among others. However, the variety and quantity of mHealth apps for HIV care and prevention are still very limited. Only 18 were available in 2018 [33], and these apps mainly focus only on 2 functionalities: (1) self-management and self-monitoring tools for increasing opioid users' adherence to medication [33-36] and (2) tools that can improve linkage and retention in HIV care among people with HIV [37,38]. None of the existing HIV-related apps have combined essential functionalities that are desired among individuals who seek HIV care and prevention, including scheduling appointments; viewing medication logs, lab reports, and current pharmacy information; tracking nutrition and fitness; and exchanging social support with other users, along with links to local resources and health information and support for self-managing stress and depression [33,37].

Study Objectives

Given the potential of mHealth apps to increase effectiveness and access to HIV care and harm reduction, we sought to better understand its feasibility and acceptance among a group of individuals who have substantial risk for HIV, that is, PWID. In addition, we sought to understand whether any participant characteristics were associated with feasibility and acceptability or with particular services that could be offered in an mHealth platform. Herein, we present and discuss our findings, focusing on the following three dimensions of feasibility and acceptability: access, comfort, and interest.

Methods

Recruitment

From June 2018 to June 2019, 115 PWID were screened for eligibility in this study at the New Haven Syringe Service Program (NHSSP). The eligibility criteria were as follows: (1) being 18 years or older; (2) being able to understand, speak, and read English or Spanish; (3) self-reporting as an active injection drug user within the past 60 days; and (4) having at least one injection partner. Respondent-driven sampling (RDS) was used for recruitment, and the original *seeds* were recruited from clientele who use the NHSSP. *Seeds* were recruited using flyers distributed at the central and mobile distribution locations of the NHSSP.

Once enrolled, *seeds* were asked to complete a cross-sectional egocentric survey in-person at the NHSSP. An iPad (Apple Inc) was used to display survey questions which were hosted on Qualtrics (SAP) to the participants, and the study staff conducting the interview selected the respondents' answers. Upon completion of the survey, each *seed* was compensated US $20 for their time. Following the interview, RDS was used to recruit subsequent waves of participants from within the *seed's* current injection network (ie, individuals who had injected drugs with the participant within the past 2 months) using the same aforementioned eligibility criteria. Participants were allowed to recruit injection partners that were either currently engaged with or not engaged with the SSP. This allowed our sample to expand beyond the limits of the SSP clientele. Each *seed* was given US $10 for each successful

XSL•FO
RenderX

network referral, and each referee was given US $20 upon completion of the survey. Once these network referees completed the survey, they were also asked to recruit their own injection network partners following the same RDS protocol. This resulted in subsequent waves of recruitment and expansion of the participant sample beyond the network of the original *seed*. All study protocols were approved by the Yale Human Subjections Committee, and a certificate of confidentiality was obtained from the National Institutes of Health to further protect participant information.

Measures

Access to Mobile Devices

We dichotomously assessed *access* to mobile devices by asking participants if they had access to a cellphone without an internet connection, a cellphone with an internet connection (smartphone), a tablet, or a computer. The main item of interest in this survey was a smartphone, which would be required for the hypothetical service app.

Comfort Levels

We assessed *comfort* levels associated with using an app and providing personal information by asking participants whether they would be willing or unwilling to provide various personal information to use the app, including their name, phone number, email, address, alternative address, and linkage to medical records. Participants also indicated their comfort level with home delivery of syringes, rated on a 5-point Likert scale, with 1 being *very comfortable* and 5 being *very uncomfortable*; a grouped outcome variable for this item was created in which *very comfortable* and *comfortable* responses were coded as 1 and all other responses were coded as 0. A summary score for *comfort* was created by adding up participant responses for each item, thus transforming it to a 0-7 scale, with higher scores indicating a higher level of comfort.

Interest

We assessed interest in potential mHealth app services by asking participants to rate the usefulness of various services, including delivery of medications or syringes, scheduling appointments with providers and labs, setting medication reminders, accessing educational material about health and safe injection practices, and accessing group communication forums with peers. Participants indicated their level of perceived usefulness on a 5-point Likert scale, with 1 being *extremely useful* and 5 being *not at all useful*; a grouped variable for each service was created in which *extremely useful* and *very useful* responses were coded as 1 and all other responses were coded as 0. In addition, 3 survey items focused on the delivery of various services (pre-exposure prophylaxis [PrEP] medication, sterile syringes, and medication for opioid use disorder), and 2 survey items focused on scheduling appointments with providers or labs were combined into 2 grouped outcomes focused on either delivery or scheduling. We created a summary score for *interest* by adding up the binary scores described above, thus creating a 0-5 scale, with higher scores indicating a higher level of interest in comprehensive services.

Participant Characteristics

We collected self-reported data on demographics, health information, and behavioral history. We measured gender identity, race and ethnicity, levels of education, and housing status (stably housed vs homeless or unstably housed) categorically and later dichotomized education at the level of high school completion. We collected age data continuously and later categorized them into the following age groups: 18-34 years, 35-49 years, and ≥50 years. We measured perceived financial stability on a 10-item Likert scale, with 1 being *always worried about food, housing, and income* and 10 being *never worried about food, housing, and income*. We later dichotomized the responses for perceived financial stability at the median level. We measured the history of incarceration, syringe sharing with current injection partners, carrying of Narcan during injection, and recent engagement with the SSP in the previous 60 days dichotomously. We similarly collected dichotomous health information, including HCV and HIV status and overdose history.

Statistical Analysis

Our primary outcomes included binary *access* to smartphones and summary scores for *comfort* and *interest*. We coded the primary *access* outcome as 1 if the participant owned a smartphone and 0 otherwise, using multivariable logistic regression when evaluating correlates. We used multivariable Poisson regression to evaluate correlations with summary scores for *comfort* and *interest*. In addition, to further explore any possible association between individuals' characteristics and their *comfort* and *interest*, we used logistic regression to evaluate correlations with each individual survey item within the *comfort* and *interest* subsections. This allowed us to see which services on the mHealth app would appeal more to certain subsets of the PWID population. As we did not correct for multiple outcomes, these secondary analyses were considered hypothesis-generating. In all models, covariates with $P<.20$ in bivariate analyses were included in the final models for each outcome. We used complete case analysis, excluding observations with missing covariate or outcome data from the bivariate or multivariable models. All statistical analyses were conducted using Stata 16 (StataCorp).

Results

Participant Characteristics

Table 1 contains the original Likert scores for the items of *interest* and *comfort* with home syringe delivery. Roughly 5.2% (6/115) of participants did not answer any questions about mHealth access, comfort (6/115, 5.2%), and interest (5/115, 4.3%), and were therefore excluded from the regression models. In general, participants were in their early 40s, primarily White, male, and had completed high school; most described themselves as homeless or unstably housed, with a minority reporting being infected with HIV or HCV. Most participants reported a history of sharing syringes with their current injection partners and overdose. In addition, just above half (64/115, 55.7%) of the participants reported using the SSP within the past 60 days, and less than half (49/115, 42.6%) of the participants reported carrying Narcan when they inject (Table 2).

XSL•FO
RenderX

Table 1. Frequencies and percentages of Likert responses for component interests and comfort engaging with home syringe delivery (n=110).

Interests[a]	Extremely useful, n (%)	Very useful, n (%)	Moderately useful, n (%)	Slightly useful, n (%)	Not at all useful, n (%)
Delivery of PrEP[b] services (n=108)	26 (24.1)	32 (29.6)	35 (32.4)	5 (4.6)	10 (9.3)
Delivery of syringes	47 (42.7)	45 (40.9)	13 (11.8)	1 (0.9)	4 (3.6)
Delivery of medications for opioid use disorder	57 (51.8)	32 (29.1)	11 (10)	1 (0.9)	9 (8.2)
Linking to electronic medical record to schedule appointments with clinical provider	43 (39.1)	41 (37.3)	20 (18.2)	5 (4.6)	1 (0.9)
Being scheduled for laboratory testing (eg, HIV, hepatitis C virus, and sexually transmitted infections; n=109)	44 (40.4)	41 (37.6)	17 (15.6)	5 (4.6)	2 (1.8)
Reminders to take medication	43 (39.1)	34 (30.9)	17 (15.5)	11 (10)	5 (4.6)
Receive educational materials	33 (30)	32 (29.1)	29 (26.4)	7 (6.4)	9 (8.2)
Group communication forums	34 (30.9)	30 (27.3)	29 (26.4)	8 (7.27)	9 (8.2)
Comfort with home syringe delivery[c] (n=109)	45 (41.3)	48 (44)	2 (1.8)	8 (7.3)	6 (5.5)

[a]Unless specified as comfort rather than interest.

[b]PrEP: pre-exposure prophylaxis.

[c]This Likert scale included *very comfortable*, *comfortable*, *unsure*, *uncomfortable*, and *very uncomfortable*.

XSL•FO
RenderX

Table 2. Characteristics and survey responses of study participants (N=115).

Participant characteristics	Values
Demographic characteristics	
Age (years)	
Value, mean (SD)	41.7 (10.6)
Age group, n (%)	
18-34	37 (32.2)
35-49	48 (41.7)
≥50	30 (26.1)
Sex, n (%)	
Male	78 (67.8)
Hispanic ethnicity, n (%)	27 (23.5)
Race (n=105), n (%)	
White	74 (70.5)
Black or African American	17 (16.2)
Other	14 (13.3)
Financial stability score, median (Q1, Q3)	3 (1, 5)
Completed high school, n (%)	87 (75.7)
Currently homeless or unstably housed, n (%)	72 (62.6)
Ever incarcerated, n (%)	96 (83.5)
Health status and behaviors	
Years of injecting, mean (SD)	14 (11.9)
Self-reported HIV positivity (n=110), n (%)	9 (8.2)
Self-reported hepatitis C virus positivity, n (%)	35 (30.4)
Syringe sharing with current injection partners, n (%)	101 (87.8)
Carry Narcan while injecting, n (%)	49 (42.6)
Recent syringe service program use, n (%)	64 (55.7)
Overdose, ever, n (%)	70 (60.9)
Mobile health opportunities (n=109), n (%)	
Access to	
Cellphone without internet	12 (11)
Smartphone	74 (67.9)
Tablet	2 (1.8)
Computer	6 (5.5)
None	22 (20)
Comfort with sharing personal information	
Name	96 (88.1)
Personal phone number	92 (84.4)
Personal email	85 (78)
Home address	85 (78)
Alternative address	62 (56.9)
Medical records	72 (66.1)
Comfort with home delivery of syringes[a]	93 (85.3)

XSL•FO
RenderX

Participant characteristics	Values
Component interests for an app[a] (n=110)	
Delivery (PrEP[b], syringes, and medications for opioid use disorder)	100 (90.9)
Clinical scheduling (provider and lab)	90 (81.8)
Medication reminders	77 (70)
Health education	65 (59.1)
Group communication forums	64 (58.2)
Mobile health summary scores for access, comfort, and interest (n=109)	
Access to smartphone, n (%)	74 (67.9)
Comfort (0-7), mean (SD)	5.4 (1.8)
Interest (0-5; n=110), mean (SD)	3.6 (1.5)

[a]See Table 1 for full Likert score responses. The proportion of participants who perceived home delivery of syringes as *comfortable* or *very comfortable* and the proportion of those who perceived services as *very* or *extremely useful* are shown here.

[b]PrEP: pre-exposure prophylaxis.

Access

Most participants owned a smartphone (74/109, 67.9%). As shown in Table 3, age, perceived financial stability, and homelessness or unstable housing reached $P<.20$ in the bivariate regression and were included in the multivariable model. In the final model, those who were homeless or unstably housed had significantly lower odds of smartphone ownership than those who were stably housed (adjusted odds ratio [AOR] 0.15, 95% CI 0.05-0.46; $P=.001$). Bivariate analyses are presented in Multimedia Appendix 1.

Table 3. Multivariable logistic and Poisson models for primary outcomes[a,b].

Covariates	Primary Outcomes					
	Access to smartphone		Comfort providing personal identifiers and engaging with personalized services		Interest in comprehensive mobile health services	
	AOR[c] (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Age (years)						
18-34	Ref[d]	Ref	—[e]	—	Ref	Ref
35-49	2.78 (0.99 to 7.82)	.05	—	—	0.05 (−0.19 to 0.30)	.68
≥50	1.04 (0.32 to 3.43)	.95	—	—	0.21 (−0.07 to 0.48)	.14
Financial stability >3	1.74 (0.66 to 4.59)	.27	—	—	—	—
Completed high school	—	—	0.13 (−0.07 to 0.33)	.19	—	—
Currently homeless or unstably housed [f]	*0.15 (0.05 to 0.46)*	*.001*	—	—	—	—
HIV positivity	—	—	—	—	0.15 (−0.21 to 0.51)	.42

[a]Logistic regression used for *access* primary outcome.

[b]Poisson regression used for *comfort* and *interest* primary outcomes.

[c]AOR: adjusted odds ratio.

[d]Ref: reference group.

[e]Not available. The covariate was not included in the final model because it did not meet the bivariate threshold of $P<.20$.

[f]Italicized text denotes significance ($P<.05$).

Comfort

Overall, most participants reported being comfortable with providing their names, phone numbers, email, address, access to medical records, and an alternative address such as a post office box on the mHealth app. In addition, a majority of participants reported being comfortable with using the home delivery of syringes on the app. Bivariate Poisson regressions showed that only education was associated with summary *comfort* scores with $P<.02$, although it was not statistically significant. No additional covariates were included in the final Poisson model.

XSL•FO
RenderX

Interest

Most participants showed interest in the proposed app features. Specifically, delivery services were perceived as at least *very useful* by most participants, as were scheduling services, setting medication reminders, accessing educational materials related to health and safe injection practices, and accessing group communication and support forums. Table 1 highlights the additional percentage of respondents that indicated only a moderate or slight interest in each service. Age and HIV status were associated with overall *interest* summary scores in bivariate Poisson regression with $P<.20$ but were not significant in the final multivariable model.

Secondary Analyses: Individual Comfort and Interest Item Analysis

Tables 4 and 5 present the multivariable findings from the secondary hypothesis-generating regression analyses. The bivariate results are presented in Multimedia Appendices 2 and 3. A few of these analyses revealed significant correlations. The middle-aged group (35-49 years) had higher odds of reporting comfort with providing their email than those who were younger (18-34 years; AOR 4.06, 95% CI 1.14-14.51; $P=.03$). In addition, those who completed high school had higher odds of reporting comfort with providing their medical records on the app than those with less education (AOR 2.87, 95% CI 1.12-7.39; $P=.03$). Those who engaged with the SSP at least once in the past 60 days had higher odds of reporting comfort with the idea of a doorstep-styled delivery of syringes (AOR 3.29, 95% CI 1.04-10.34; $P=.04$). Older PWID (≥50 years) had higher odds of reporting interest in having educational materials (AOR 4.64, 95% CI 1.31-16.46; $P=.02$) and communication forums on the app (AOR 3.69, 95% CI 1.10-12.44; $P=.04$). Those who had been previously incarcerated had lower odds of reporting interest in educational materials on the app (AOR 0.22, 95% CI 0.05-0.87; $P=.03$).

XSL•FO
RenderX

Table 4. Secondary multivariable logistic regression models evaluating individual comfort items.

Covariates	Individual comfort items											
	Name		Email		Phone number		Address		Medical records		Home syringe delivery	
	AOR[a] (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value
Age (years)												
18-34	Ref[b]	Ref	Ref	Ref	—[c]	—	Ref	Ref	—	—	—	—
35-49 [d]	0.51 (0.09-2.92)	.45	*4.06 (1.14-14.51)*	*.03*	—	—	2.26 (0.75-6.78)	.15	—	—	—	—
≥50	0.49 (0.05-5.23)	.56	0.77 (0.25-2.39)	.65	—	—	1.22 (0.39-3.81)	.74	—	—	—	—
Female	—	—	—	—	—	—	2.05 (0.69-6.13)	.20	—	—	—	—
Hispanic ethnicity	—	—	0.36 (0.11-1.12)	.08	—	—	—	—	—	—	—	—
Completed high school	1.74 (0.48-6.26)	.40	—	—	—	—	—	—	*2.87 (1.12-7.39)*	*.03*	—	—
Homeless or unstably housed	0.20 (0.39-1.00)	.05	—	—	—	—	—	—	0.52 (0.22-1.27)	.15	—	—
Ever been incarcerated	—	—	2.50 (0.78-8.10)	.13	—	—	—	—	—	—	—	—
Years of injecting	0.96 (0.90-1.03)	.29	—	—	—	—	—	—	0.98 (0.95-1.02)	.37	—	—
Recent SSP use [e]	—	—	—	—	1.79 (0.61-5.24)	.29	—	—	—	—	*3.29 (1.04-10.34)*	*.04*
Carry Narcan	—	—	—	—	1.96 (0.62-6.15)	.25	—	—	—	—	—	—
Syringe sharing	—	—	—	—	—	—	—	—	2.21 (0.67-7.26)	.19	2.69 (0.70-10.37)	.15

[a]AOR: adjusted odds ratio.

[b]Ref: reference group.

[c]Not available. The covariate was not included in the final model because it did not meet the bivariate threshold of $P<.20$.

[d]Italicized text denotes significance ($P<.05$).

[e]SSP: syringe service program.

XSL•FO
RenderX

Table 5. Secondary multivariable logistic regression models evaluating individual interest items.

Covariates	Individual interest items									
	Delivery		Scheduling		Reminders		Educational material		Communication forums	
	AOR[a] (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value	AOR (95% CI)	*P* value
Age (years)										
18-34	—[b]	—	—	—	Ref[c]	Ref	Ref	Ref	Ref	Ref
35-49	—	—	—	—	1.79 (0.65-4.91)	.26	0.97 (0.36-2.59)	.95	1.23 (0.46-3.29)	.68
≥50[d]	—	—	—	—	5.05 (0.95-26.72)	.06	*4.64 (1.31-16.46)*	*.02*	*3.69 (1.10-12.44)*	*.04*
Race										
White	Ref	Ref	—	—	Ref	Ref	—	—	Ref	Ref
Black or African American	0.37 (0.07-1.96)	.24	—	—	1.59 (0.38-6.65)	.52	—	—	1.31 (0.37-4.63)	.67
Other	0.28 (0.04-1.91)	.20	—	—	1.15 (0.34-3.97)	.82	—	—	1.13 (0.32-4.02)	.85
Female	2.90 (0.32-26.86)	.35	—	—	—	—	—	—	—	—
Hispanic ethnicity	—	—	0.44 (0.14-1.33)	.15	—	—	2.17 (0.69-6.75)	.18	2.01 (0.57-7.02)	.28
Completed high school	—	—	1.88 (0.62-5.70)	.27	—	—	—	—	—	—
Ever been incarcerated	—	—	—	—	—	—	*0.22 (0.05-0.87)*	*.03*	—	—
HIV positivity	—	—	—	—	—	—	3.52 (0.36-34.62)	.28	—	—
HCV[e] positivity	—	—	—	—	—	—	—	—	0.46 (0.19-1.14)	.09
Years of injecting	0.97 (0.92-1.03)	.34	—	—	1.00 (0.95-1.05)	.99	—	—	—	—
Carry Narcan	3.29 (0.61-17.73)	.17	—	—	—	—	0.57 (0.24-1.38)	.21	—	—

[a]AOR: adjusted odds ratio.

[b]Not available. The covariate was not included in the final model because it did not meet the bivariate threshold of $P<.20$.

[c]Ref: reference group.

[d]Italicized text denotes significance ($P<.05$).

[e]HCV: hepatitis C virus.

Discussion

Principal Findings

Our results demonstrate a high prevalence of smartphone ownership and diffuse comfort and interest associated with mHealth services among the PWID included in this study. The results of our primary analyses were largely null aside from finding that those that were homeless or unstably housed had lower odds of smartphone ownership. In combination with the high rates of reported comfort and interest, this suggests that such an app-based approach would be acceptable for a wide range of PWID. The hypothetical services included in our survey would each offer unique methods of increasing harm reduction service uptake in the PWID community by overcoming various barriers to SSP engagement and warrant further research and development in this area.

Few significant correlations were found in our primary regression analyses, indicating that the acceptability of an mHealth app is diffusely similar across subsets of PWID. Although most participants reported low financial stability, most owned smartphones and also reported being comfortable with providing their personal information such as name, phone number, email, and address to be able to use an mHealth app. The smartphone ownership level among participants was high but lower than the 81% reported nationally in the United States [39]. Smartphone ownership was lower among those

XSL•FO
RenderX

experiencing homelessness or unstable housing, indicating that reaching this group with an mHealth app could be challenging. The comfort associated with providing linkage to medical records was slightly lower than the other comfort survey items, indicating that integrating electronic health records in the app could be a barrier for those who have privacy concerns. This may more commonly include those with lower levels of education, as suggested by our secondary analyses. Nonetheless, two-thirds of the participants were comfortable with providing access to medical records and could benefit from linking harm reduction to health care services. Most participants showed a strong desire for a contact-free format to obtain syringes and medications via the app. This service could address current barriers (eg, limited hours and personnel and stigmatization) associated with the current in-person, face-to-face format of most SSPs. Overall, the mostly null findings from our primary analyses indicate a widespread acceptability of an mHealth app designed for PWID.

Our secondary hypothesis-generating analyses revealed significant associations between age and interest in educational materials and group discussion forums and comfort with providing access to personal email. We also found education to be associated with comfort with providing linkage to medical records and a history of incarceration to be associated with interest in educational materials. Although only hypothesis-generating, these secondary findings indicate that allowing customization in an app is crucial for a positive user experience, especially for those with particular interests or privacy concerns. Of particular importance is the secondary finding that recent SSP use increased the odds of comfort with the home delivery of syringes. Aligned with the theory of diffusion by Roger [40,41], this finding indicates that PWID who are currently engaged with SSPs are likely to be the key influencers of the diffusion of an mHealth app and may drive the initial adoption of an mHealth app because of their established trust and familiarity with harm reduction providers and services. The subsequent diffusion of the app to PWID not currently engaged with SSPs may rely on early adopters to encourage their social network contacts to adopt the app.

The variety and quantity of mHealth apps designed for PWID and HIV or HCV care and prevention are very limited [33,42]. Most current apps serve as self-monitoring tools to increase adherence to medication among opioid users or people with HIV [33-36]. To our knowledge, there are no mHealth apps that use a patient-to-provider model offering broad and comprehensive services for PWID, which would greatly advance the mHealth field given the high levels of medical and psychiatric comorbidity in this population [43]. Our findings indicate that a multifunctional mHealth app that can deliver medication reminders, educational materials, and communication forums is highly desired by PWID, the latter especially for older PWID. This app may help reduce the risk of adverse health outcomes and build upon the core harm reduction services such as medication (eg, antiretroviral therapy, PrEP, HCV antivirals, and buprenorphine) and syringe provision [33]. Providing medication reminders for PWID, especially tailored messages, on an app may improve their self-management and medication adherence, which is particularly critical for PWID taking PrEP or antiretroviral therapy [44,45]. In addition, educational materials would also offer PWID information about how to reduce their risks while injecting and educate about overdose prevention and naloxone use. As drug injection is an illicit and often taboo activity, PWID must carefully navigate their environments while seeking information about or assistance with safe injection practices. This is particularly the case for those who have recently transitioned to drug injection, a growing subgroup of PWID that experience a particularly high risk of adverse outcomes, especially overdose, because of gaps in health knowledge, inexperience, and risk behaviors [46]. Our study extends this literature by showing that such education is also important for older PWID, and an mHealth app may be an effective, highly preferred platform for delivering educational material. By providing such educational material within an app, PWID can deploy safe injection practices using instructions within their reach. Finally, our study shows that communication forums may also be favored by older PWID. This demonstrates the older PWID's desire for web-based communication and support. Such communication forums with peers may offer PWID a means to exchange information on safe injection practices and lessons learned and also provide encouragement and emotional support in a safe, monitored environment.

Limitations

Despite the many new findings, this study has several limitations. This study included a relatively small sample size, which may have resulted in type II errors, and the fact that all initial *seed* PWID were enrolled from the SSP clientele may have resulted in sampling bias. Previous evaluation of RDS methods suggest that 4-6 recruitment waves are needed to overcome sampling bias with regard to race, sex, and drug use status [47], but the mean number of recruitment waves per *seed* in our study was 1.59. However, the most productive network in this study consisted of 12 waves and produced 56.5% (65/115) of the study participants, and a total of 83.5% (96/115) of our sample stemmed from seeds with at least three recruitment waves, suggesting that we were able to overcome some, but not all, of the sampling bias.

Strengths of This Work

A strength of our study was our ability to recruit and interview PWID who do not currently engage with the local SSP; 46.1% (53/115) of the participants were not currently engaged with the NHSSP. These participants represent a large and hidden population of PWID that is not reached by the current brick-and-mortar SSP models. Their responses regarding their preferences and barriers to using SSPs give us crucial insight into what improvements are needed to increase the accessibility of SSPs to better address the health needs of this population, including through mHealth delivery strategies. Furthermore, this study provided insight into a broad range of relevant issues regarding mHealth feasibility and interest, including access to required devices, comfort levels, and interest.

Conclusions

In conclusion, our findings suggest high acceptability of an mHealth app among PWID with little differences between subgroups of PWID, suggesting that an app would not only be

XSL•FO
RenderX

applicable to a niche of PWID. The barriers to traditional SSP engagement, including awareness, access, and stigma, may be circumvented with such a novel intervention, and offering alternative solutions via a mobile platform could greatly increase the reach of SSPs and reduce PWIDs' risk of harm significantly. Further research is warranted on this topic, and in-depth interviews and focus-group discussions should be used to provide more detailed information about how such an app could best serve the target audience.

Acknowledgments

The authors would like to acknowledge the tireless efforts of the staff at the NHSSP, including Sharon Joslin, Angel Ojeda, Migdalia Williams, Rodolfo Lopez, Rodolfo Lopez Jr, Barbara Valdes, Carolina Price, Lisandra Alvarez, Constance Carroll, and Paula Dellamura. Without their critical insight, fruitful conversations, and endless support, this work would not have been possible. Finally, the authors would like to thank Alexei Zelenev for his helpful discussions and methodological guidance. During the study period, TS was supported by a T32 training grant, and XZ was supported by an R21 grant. Additional funding support for this work was provided by R21 DA039842, R01 DA033679, and an independent sponsored research grant from Merck.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bivariate logistic and Poisson models for primary outcomes.
[DOCX File , 21 KB - jmir_v23i7e25428_app1.docx]

Multimedia Appendix 2
Secondary bivariate logistic regression models evaluating individual comfort items.
[DOCX File , 24 KB - jmir_v23i7e25428_app2.docx]

Multimedia Appendix 3
Secondary bivariate logistic regression models evaluating individual interest items.
[DOCX File , 23 KB - jmir_v23i7e25428_app3.docx]

References

1. Fleischauer AT, Ruhl L, Rhea S, Barnes E. Hospitalizations for endocarditis and associated health care costs among persons with diagnosed drug dependence - North Carolina, 2010-2015. MMWR Morb Mortal Wkly Rep 2017 Jun 09;66(22):569-573 [FREE Full text] [doi: 10.15585/mmwr.mm6622a1] [Medline: 28594786]
2. Schranz AJ, Barrett J, Hurt CB, Malvestutto C, Miller WC. Challenges facing a rural opioid epidemic: treatment and prevention of HIV and Hepatitis C. Curr HIV/AIDS Rep 2018 Jun;15(3):245-254 [FREE Full text] [doi: 10.1007/s11904-018-0393-0] [Medline: 29796965]
3. Ciccarone D, Unick GJ, Cohen JK, Mars SG, Rosenblum D. Nationwide increase in hospitalizations for heroin-related soft tissue infections: associations with structural market conditions. Drug Alcohol Depend 2016 Jun 01;163:126-133 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.04.009] [Medline: 27155756]
4. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, et al. Increases in acute Hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. Am J Public Health 2018 Feb;108(2):175-181. [doi: 10.2105/AJPH.2017.304132] [Medline: 29267061]
5. Schwetz TA, Calder T, Rosenthal E, Kattakuzhy S, Fauci AS. Opioids and infectious diseases: a converging public health crisis. J Infect Dis 2019 Jul 02;220(3):346-349 [FREE Full text] [doi: 10.1093/infdis/jiz133] [Medline: 30941402]
6. Opioid overdose: understanding the epidemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/drugoverdose/epidemic/index.html [accessed 2021-06-02]
7. Aspinall EJ, Nambiar D, Goldberg DJ, Hickman M, Weir A, Van VE, et al. Are needle and syringe programmes associated with a reduction in HIV transmission among people who inject drugs: a systematic review and meta-analysis. Int J Epidemiol 2014 Feb;43(1):235-248 [FREE Full text] [doi: 10.1093/ije/dyt243] [Medline: 24374889]
8. Fernandes RM, Cary M, Duarte G, Jesus G, Alarcão J, Torre C, et al. Effectiveness of needle and syringe programmes in people who inject drugs - an overview of systematic reviews. BMC Public Health 2017 Apr 11;17(1):309 [FREE Full text] [doi: 10.1186/s12889-017-4210-2] [Medline: 28399843]
9. Grau LE, Griffiths-Kundishora A, Heimer R, Hutcheson M, Nunn A, Towey C, et al. Barriers and facilitators of the HIV care continuum in Southern New England for people with drug or alcohol use and living with HIV/AIDS: perspectives of HIV surveillance experts and service providers. Addict Sci Clin Pract 2017 Oct 02;12(1):24 [FREE Full text] [doi: 10.1186/s13722-017-0088-7] [Medline: 28965489]

XSL•FO
RenderX

10. Altice FL, Springer S, Buitrago M, Hunt DP, Friedland GH. Pilot study to enhance HIV care using needle exchange-based health services for out-of-treatment injecting drug users. J Urban Health 2003 Sep;80(3):416-427 [FREE Full text] [doi: 10.1093/jurban/jtg053] [Medline: 12930880]
11. Heimer R, Grau LE, Curtin E, Khoshnood K, Singer M. Assessment of HIV testing of urban injection drug users: implications for expansion of HIV testing and prevention efforts. Am J Public Health 2007 Jan;97(1):110-116. [doi: 10.2105/AJPH.2005.078105] [Medline: 17138916]
12. Abbasi J. CDC says more needle exchange programs needed to prevent HIV. J Am Med Assoc 2017 Jan 24;317(4):350. [doi: 10.1001/jama.2016.19452] [Medline: 28118437]
13. Rockwell R, Des JD, Friedman SR, Perlis TE, Paone D. Geographic proximity, policy and utilization of syringe exchange programmes. AIDS Care 1999 Aug;11(4):437-442. [doi: 10.1080/09540129947811] [Medline: 10533536]
14. Williams CT, Metzger DS. Race and distance effects on regular syringe exchange program use and injection risks: a geobehavioral analysis. Am J Public Health 2010 Jun;100(6):1068-1074 [FREE Full text] [doi: 10.2105/AJPH.2008.158337] [Medline: 20395589]
15. Canary L, Hariri S, Campbell C, Young R, Whitcomb J, Kaufman H, et al. Geographic disparities in access to syringe services programs among young persons with Hepatitis C virus infection in the United States. Clin Infect Dis 2017 Aug 01;65(3):514-517. [doi: 10.1093/cid/cix333] [Medline: 28402431]
16. Sticking points: barriers to access to needle and syringe programs in Canada. HIV Legal Network. 2007. URL: https://tinyurl.com/jf5prhsp [accessed 2021-06-16]
17. Strathdee SA, Ricketts EP, Huettner S, Cornelius L, Bishai D, Havens JR, et al. Facilitating entry into drug treatment among injection drug users referred from a needle exchange program: Results from a community-based behavioral intervention trial. Drug Alcohol Depend 2006 Jul 27;83(3):225-232 [FREE Full text] [doi: 10.1016/j.drugalcdep.2005.11.015] [Medline: 16364566]
18. Aronson ID, Bennett A, Marsch LA, Bania TC. Mobile technology to increase HIV/HCV testing and overdose prevention/response among people who inject drugs. Front Public Health 2017;5:217 [FREE Full text] [doi: 10.3389/fpubh.2017.00217] [Medline: 28879174]
19. Davis SM, Kristjansson AL, Davidov D, Zullig K, Baus A, Fisher M. Barriers to using new needles encountered by rural Appalachian people who inject drugs: implications for needle exchange. Harm Reduct J 2019 Apr 02;16(1):23 [FREE Full text] [doi: 10.1186/s12954-019-0295-5] [Medline: 30940136]
20. Naserirad M, Beulaygue IC. Accessibility of needle and syringe programs and injecting and sharing risk behaviors in high Hepatitis C virus prevalence settings. Subst Use Misuse 2020;55(6):900-908. [doi: 10.1080/10826084.2019.1710210] [Medline: 31918611]
21. Paquette CE, Syvertsen JL, Pollini RA. Stigma at every turn: health services experiences among people who inject drugs. Int J Drug Policy 2018 Jul;57:104-110 [FREE Full text] [doi: 10.1016/j.drugpo.2018.04.004] [Medline: 29715589]
22. Rivera AV, DeCuir J, Crawford ND, Amesty S, Lewis CF. Internalized stigma and sterile syringe use among people who inject drugs in New York City, 2010-2012. Drug Alcohol Depend 2014 Nov 01;144:259-264 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.09.778] [Medline: 25307745]
23. Sherman SG, Patel SA, Ramachandran DV, Galai N, Chaulk P, Serio-Chapman C, et al. Consequences of a restrictive syringe exchange policy on utilisation patterns of a syringe exchange program in Baltimore, Maryland: implications for HIV risk. Drug Alcohol Rev 2015 Nov;34(6):637-644 [FREE Full text] [doi: 10.1111/dar.12276] [Medline: 25919590]
24. Strike C, Miskovic M. Scoping out the literature on mobile needle and syringe programs-review of service delivery and client characteristics, operation, utilization, referrals, and impact. Harm Reduct J 2018 Feb 08;15(1):6 [FREE Full text] [doi: 10.1186/s12954-018-0212-3] [Medline: 29422042]
25. Strike CJ, Challacombe L, Myers T, Millson M. Needle exchange programs. Delivery and access issues. Can J Public Health 2002;93(5):339-343 [FREE Full text] [Medline: 12353453]
26. Shrestha R, Huedo-Medina TB, Altice FL, Krishnan A, Copenhaver M. Examining the acceptability of mHealth technology in HIV prevention among high-risk drug users in treatment. AIDS Behav 2017 Nov;21(11):3100-3110 [FREE Full text] [doi: 10.1007/s10461-016-1637-x] [Medline: 28025735]
27. Alloghani M, Hussain A, Al-Jumeily D, Fergus P, Abuelma'Atti O, Hamden H. A mobile health monitoring application for obesity management and control using the internet-of-things. In: Proceedings of the Sixth International Conference on Digital Information Processing and Communications (ICDIPC). 2016 Presented at: Sixth International Conference on Digital Information Processing and Communications (ICDIPC); April 21-23, 2016; Beirut, Lebanon. [doi: 10.1109/icdipc.2016.7470785]
28. Al AS, Parmanto B, Branch R, Ding D. A persuasive and social mhealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
29. Hartzler A, Venkatakrishnan A, Mohan S, Silva M, Lozano P, Ralston J. Acceptability of a team-based mobile health (mHealth) application for lifestyle self-management in individuals with chronic illnesses. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug. 16-20, 2016; Orlando, FL, USA. [doi: 10.1109/embc.2016.7591428]

XSL•FO
RenderX

30. García L, Tomás J, Parra L, Lloret J. An m-health application for cerebral stroke detection and monitoring using cloud services. Int J Inf Manage 2019 Apr;45:319-327. [doi: 10.1016/j.ijinfomgt.2018.06.004]
31. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
32. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
33. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Jan 03:3373-3383. [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
34. Guarino H, Acosta M, Marsch LA, Xie H, Aponte-Melendez Y. A mixed-methods evaluation of the feasibility, acceptability, and preliminary efficacy of a mobile intervention for methadone maintenance clients. Psychol Addict Behav 2016 Feb;30(1):1-11 [FREE Full text] [doi: 10.1037/adb0000128] [Medline: 26618796]
35. Himelhoch S, Kreyenbuhl J, Palmer-Bacon J, Chu M, Brown C, Potts W. Pilot feasibility study of Heart2HAART: a smartphone application to assist with adherence among substance users living with HIV. AIDS Care 2017 Dec;29(7):898-904. [doi: 10.1080/09540121.2016.1259454] [Medline: 28043176]
36. Perera AI, Thomas MG, Moore JO, Faasse K, Petrie KJ. Effect of a smartphone application incorporating personalized health-related imagery on adherence to antiretroviral therapy: a randomized clinical trial. AIDS Patient Care STDS 2014 Nov;28(11):579-586 [FREE Full text] [doi: 10.1089/apc.2014.0156] [Medline: 25290556]
37. Laurence C, Wispelwey E, Flickinger TE, Grabowski M, Waldman AL, Plews-Ogan E, et al. Development of PositiveLinks: a mobile phone app to promote linkage and retention in care for people with HIV. JMIR Form Res 2019 Mar 20;3(1):e11578 [FREE Full text] [doi: 10.2196/11578] [Medline: 30892269]
38. Westergaard RP, Genz A, Panico K, Surkan PJ, Keruly J, Hutton HE, et al. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addict Sci Clin Pract 2017 Dec 26;12(1):11 [FREE Full text] [doi: 10.1186/s13722-017-0076-y] [Medline: 28441962]
39. Mobile fact sheet 2019. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-06-10]
40. Rogers E. Diffusion of innovations. New York, United States: Simon and Schuster; 2010:1-576.
41. Glanz K, Lewis FM, Rimer BK. Health Behavior and Health Education. Medicine & Science in Sports & Exercise 1991;23(12):1404. [doi: 10.1249/00005768-199112000-00016]
42. Rodrigues AT, Sousa CT, Pereira J, Figueiredo IV, Lima TD. Mobile applications (Apps) to support the Hepatitis C treatment: a systematic search in app stores. Ther Innov Regul Sci 2021 Jan;55(1):152-162. [doi: 10.1007/s43441-020-00201-8] [Medline: 32700147]
43. Altice FL, Kamarulzaman A, Soriano VV, Schechter M, Friedland GH. Treatment of medical, psychiatric, and substance-use comorbidities in people infected with HIV who use drugs. Lancet 2010 Jul 31;376(9738):367-387 [FREE Full text] [doi: 10.1016/S0140-6736(10)60829-X] [Medline: 20650518]
44. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
45. Shrestha R, Altice F, Karki P, Copenhaver M. Developing an integrated, brief biobehavioral HIV prevention intervention for high-risk drug users in treatment: the process and outcome of formative research. Front Immunol 2017;8:561 [FREE Full text] [doi: 10.3389/fimmu.2017.00561] [Medline: 28553295]
46. Jost JJ, Tempalski B, Vera T, Akiyama MJ, Mangalonzo AP, Litwin AH. Gaps in HCV knowledge and risk behaviors among young suburban people who inject drugs. Int J Environ Res Public Health 2019 Jun 02;16(11):1958 [FREE Full text] [doi: 10.3390/ijerph16111958] [Medline: 31159479]
47. Abdul-Quader AS, Heckathorn DD, McKnight C, Bramson H, Nemeth C, Sabin K, et al. Effectiveness of respondent-driven sampling for recruiting drug users in New York City: findings from a pilot study. J Urban Health 2006 May;83(3):459-476 [FREE Full text] [doi: 10.1007/s11524-006-9052-7] [Medline: 16739048]

Abbreviations

AOR: adjusted odds ratio
HCV: hepatitis C virus
mHealth: mobile health
NHSSP: New Haven Syringe Service Program
PrEP: pre-exposure prophylaxis
PWID: people who inject drugs
RDS: respondent-driven sampling
SSP: syringe service program

XSL•FO
RenderX

Edited by R Kukafka; submitted 01.11.20; peer-reviewed by J Jain, R Jeminiwa; comments to author 15.01.21; revised version received 27.01.21; accepted 24.05.21; published 14.07.21.

Please cite as:
Shelby T, Zhou X, Barber D, Altice F
Acceptability of an mHealth App That Provides Harm Reduction Services Among People Who Inject Drugs: Survey Study
J Med Internet Res 2021;23(7):e25428
URL: https://www.jmir.org/2021/7/e25428
doi:10.2196/25428
PMID:34259640

©Tyler Shelby, Xin Zhou, Douglas Barber, Frederick Altice. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Ecological Momentary Assessment Through a Passive Smartphone-Based App (eB2) by Patients With Schizophrenia: Acceptability Study

Javier-David Lopez-Morinigo[1,2,3,4], MD, PhD; María Luisa Barrigón[1,2], MD, PhD; Alejandro Porras-Segovia[1,5], MD, PhD; Verónica González Ruiz-Ruano[1,2], MD; Adela Sánchez Escribano Martínez[1,2], MD; Paula Jhoana Escobedo-Aedo[1], MD; Sergio Sánchez Alonso[1], MD; Laura Mata Iturralde[1], MD, PhD; Laura Muñoz Lorenzo[1], MD, PhD; Antonio Artés-Rodríguez[3,6,7,8], PhD; Anthony S David[9], MD; Enrique Baca-García[1,2,3,5,10,11,12,13], MD, PhD

[1]Departamento de Psiquiatria, IIS-Fundación Jiménez Díaz, Madrid, Spain

[2]Departamento de Psiquiatria, Universidad Autónoma de Madrid, Madrid, Spain

[3]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain

[4]Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, School of Medicine, Universidad Complutense, Madrid, Spain

[5]Departamento de Psiquiatria, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain

[6]Departamento de Teoría de Señal y de la Comunicación, Universidad Carlos III, Madrid, Spain

[7]Instituto de Investigación Sanitaria Gregorio Marañón (IiSGM), Madrid, Spain

[8]Evidence-Based Behavior, Leganés, Madrid, Spain

[9]Institute of Mental Health, University College London, London, United Kingdom

[10]Universidad Católica del Maule, Talca, Chile

[11]Departamento de Psiquiatría, Hospital Universitario Central de Villalba, Madrid, Spain

[12]Departamento de Psiquiatría, Hospital Universitario Infanta Elena, Valdemoro, Madrid, Spain

[13]Université de Nîmes, Nimes, France

Corresponding Author:
Javier-David Lopez-Morinigo, MD, PhD
Departamento de Psiquiatria, IIS-Fundación Jiménez Díaz
Avenida de los Reyes Católicos
Madrid, 28040
Spain
Phone: 34 627277126
Email: javierd.lopez@uam.es

Abstract

Background: Ecological momentary assessment (EMA) tools appear to be useful interventions for collecting real-time data on patients' behavior and functioning. However, concerns have been voiced regarding the acceptability of EMA among patients with schizophrenia and the factors influencing EMA acceptability.

Objective: The aim of this study was to investigate the acceptability of a passive smartphone-based EMA app, evidence-based behavior (eB2), among patients with schizophrenia spectrum disorders and the putative variables underlying their acceptance.

Methods: The participants in this study were from an ongoing randomized controlled trial (RCT) of metacognitive training, consisting of outpatients with schizophrenia spectrum disorders (F20-29 of 10th revision of the International Statistical Classification of Diseases and Related Health Problems), aged 18-64 years, none of whom received any financial compensation. Those who consented to installation of the eB2 app (users) were compared with those who did not (nonusers) in sociodemographic, clinical, premorbid adjustment, neurocognitive, psychopathological, insight, and metacognitive variables. A multivariable binary logistic regression tested the influence of the above (independent) variables on "being user versus nonuser" (acceptability), which was the main outcome measure.

Results: Out of the 77 RCT participants, 24 (31%) consented to installing eB2, which remained installed till the end of the study (median follow-up 14.50 weeks) in 14 participants (70%). Users were younger and had a higher education level, better premorbid adjustment, better executive function (according to the Trail Making Test), and higher cognitive insight levels (measured with

XSL•FO
RenderX

the Beck Cognitive Insight Scale) than nonusers (univariate analyses) although only age (OR 0.93, 95% CI 0.86-0.99; *P*=.048) and early adolescence premorbid adjustment (OR 0.75, 95% CI 0.61-0.93; *P*=.01) survived the multivariable regression model, thus predicting eB2 acceptability.

Conclusions: Acceptability of a passive smartphone-based EMA app among participants with schizophrenia spectrum disorders in this RCT where no participant received financial compensation was, as expected, relatively low, and linked with being young and good premorbid adjustment. Further research should examine how to increase EMA acceptability in patients with schizophrenia spectrum disorders, in particular, older participants and those with poor premorbid adjustment.

Trial Registration: ClinicalTrials.gov NCT04104347; https://clinicaltrials.gov/ct2/show/NCT04104347

(J Med Internet Res 2021;23(7):e26548) doi:10.2196/26548

KEYWORDS

ecological momentary assessment; acceptability; schizophrenia spectrum disorders; eB2; digital tools; mental health; schizophrenia; real-time data; patients; digital health; internet; mobile apps

Introduction

Up to 4.1 billion people were reported using the internet in 2019 and 83% of the population worldwide reported having a mobile broadband subscription with internet access, although there were considerable differences between low-income and high-income countries [1]. e-Mental health has become an emerging field for a wide range of mental disorders [2], with a significantly increased number of e-mental health papers published since 1993 [3]. In particular, ecological momentary assessment (EMA) via smartphones (actively or passively) appears to be a clinically useful resource, which enables clinicians and researchers to build a digital phenotype. Digital phenotyping, defined as "the moment-by-moment quantification of the individual-level human phenotype in situ by using data from personal digital devices," [4] has been applied to a variety of mental disorders [5,6]. In particular, passive EMA, which does not require any active role by the participant, provides 2 significant advantages over traditional data collection methods through validated scales or questionnaires. First, real-time data recording avoids recall bias to which patients are subjected at the time of being interviewed. Second, follow-up face-to-face assessments are not needed, which reduce the risk of attrition and related issues [7].

Schizophrenia can be considered as the most severe mental illness, given its poor clinical and social outcomes [8] and excess mortality [9]. Although EMA has been demonstrated to be a valid and feasible resource in a wide range of mental disorders, including schizophrenia spectrum disorders (SSD) [2], concerns have been raised regarding its safety and acceptability among individuals with SSD. In particular, EMA may exacerbate paranoid thoughts [10], although this remains unclear, and to properly research this, may raise ethical issues [4]. Indeed, concerns about acceptability of EMA methods among patients with SSD have been raised, particularly taking into account the overall poor compliance in patients with SSD [11], which may, however, be improved by EMA-based interventions. In keeping with this, data from a 2016 meta-analysis [12] revealed that up to 66.4% of people with psychosis own a mobile phone, which is likely to have increased by now. Over a decade ago, high levels of acceptance (96%) and compliance (87%) with computerized laboratory methods in patients with schizophrenia and schizoaffective disorders were found, with financial compensation (from US $35 to US $100) provided to participants [13]. Consistent with this, acceptability levels of EMA methods among patients with SSD were relatively good-to-excellent in terms of retention, that is, 92%, according to a 2015 meta-analysis of 5 studies on smartphone-based EMA tools and follow-up periods ranging from 6 to 130 days [14]. However, participants in some of the selected studies received a financial incentive, which may have incorporated a selection bias. In addition, cognitive impairment [15], cannabis use [16], negative symptom severity [17], and clinical appointment attendance [18] were linked with poorer EMA acceptability and compliance in patients with SSD, and illness severity and relapse may lead to exclusion from studies and attrition [19,20]. However, passive smartphone-based EMA apps, which require no cooperation from the subject [21], may increase acceptability. Thus, passive EMA metrics, in addition to either active EMA or self-report scales, were reported to be a feasible tool in patients with SSD [20], which can be useful in monitoring clinical improvement during the acute inpatient episode [22] and in assessing a range of illness aspects such as sleep disturbances [23], autonomic deregulation [24], positive and negative emotions [25], symptom severity [21], social stress [26], and functioning [27]. In addition, EMA may aid in predicting clinical outcomes such as transition from "at-high-risk for psychosis" to having a psychotic disorder [28] and relapsing [29]. EMA-based tools such as reminders via text messages may also have a role in treatment [30]. Indeed, a recent meta-analysis, which included 9 studies, showed that EMA improved symptom severity and compliance [19]. Thus, further replication studies testing passive EMA devices in isolation are warranted.

Two previous investigations by our group examined EMA use, acceptability, and compliance, none of which focused on patients with SSD. The first study examined the characteristics of mental health service users who actively used active web-based EMA methods such as MeMind [31] over a 1-year follow-up. Interestingly, out of 13,811 subjects who were registered for MeMind, over 20% of them (2838/13,811) used the active interface on at least one occasion [2]. Later on, a case-control study was designed to test the acceptability of the aforementioned MeMind, which is active (ie, it requires user's collaboration), and a passive smartphone-based app called the evidence-based behavior (eB2) among patients with mental

XSL•FO
RenderX

health disorders with/without a history of suicidal thoughts and behavior and in healthy students (controls) [32]. Regarding the eB2 app, acceptability levels among patients receiving mental health care ranged from 71.7% to 73.5%, although at 2-month follow-up, retention rates dropped to approximately 65% [32].

By building on this work, we aimed to investigate the acceptability of a passive EMA smartphone-based app, eB2, among patients with SSD and what variables predicted this acceptability. Of note, we did not aim to investigate the proportion of patients with SSD who continued using eB2 or the factors influencing the usage of this EMA app. Rather, our research question focused on the extent to which patients with SSD gave consent to installation of a smartphone-based app, which was immediately downloaded by a researcher (VGRR) in front of the patient at that point (ie, acceptability) and the factors underlying this (ie, putative predictors of eB2 acceptability). Specifically, 3 hypotheses were tested. Based on the aforementioned study [32], we postulated that (1) the proportion of randomized controlled trial (RCT) participants consenting to eB2 will be lower than 50% (ie, low recruitment levels) and (2) only a small proportion of eB2 users will continue using the app till the end of the study period (ie, low retention). In addition, we hypothesized that eB2 users (compared with nonusers) will have better neurocognitive and metacognitive performance, less severe psychopathological symptoms, and greater insight levels, thus emerging as predictors of EMA acceptability among patients with SSD and, somewhat, replicating previous findings from our group regarding the acceptability of MeMind [2].

Methods

Sample Population

The participants in this study were from an ongoing RCT of metacognitive training, which is being carried out at the Hospital Universitario Fundación Jiménez Díaz (Madrid, Spain) [33]. Briefly, these outpatients (age 18-64 years) with an SSD diagnosis (F20-29 of 10th revision of the International Statistical Classification of Diseases and Related Health Problems), according to the Mini International Neuropsychiatric Interview, 5th edition [34], from June 10, 2019 to March 11, 2020, were considered to be eligible. Recruitment had to be stopped on March 11, 2020 due to the COVID-19 outbreak in Spain. Exclusion criteria were (1) an intelligence quotient ≤70, which was assessed with the short form of the Wechsler Adults Intelligence Scale-IV [35]; (2) a history of head injury or a neurological condition; (3) having received a metacognitive intervention within the previous year; (4) low level of communication in Spanish; and (5) clinician judgment of the participant being unable to complete all aspects of the RCT (eg, clinician-perceived cognitive difficulty of the patient to complete all assessment procedures or weekly therapy group sessions). Of relevance, eligible candidates were reassured that refusing to participate in or dropping out of the study at any time would have no implications on care provision. While customary in research projects, financial compensation of participants may prevent us from fully understanding the acceptability of EMA apps in real-world conditions. In particular, it should be noted that the vast majority of patients with mental health problems receive publicly funded care in Spain, which is free at the point of delivery. Although participants may have been financially compensated for their time, we considered that this may have affected the external validity of our results in terms of EMA acceptability, particularly in our setting. This RCT obtained ethical approval from the local research ethics committee and is registered at clinicaltrials.gov (NCT04104347). Participants gave written informed consent to the research project principal investigator (JDLM) who led the first face-to-face interview with eligible candidates. Those who agreed to participate in the trial were asked if they owned a smartphone. If this was the case, consent to eB2 installation (which is explained below) was attempted to be obtained. Another researcher (VGRR) provided participants in the trial with all the relevant information on eB2, particularly regarding their passive role. More specifically, participants were explained that after installation of the app by one researcher (VGRR) at the clinic in front of them, they did not have to upload data. Equally, they were informed about how to uninstall the app at any time and were provided with the team telephone number. Those who consented to eB2 (users) and those who did not (nonusers) were compared on sociodemographic, clinical, neurocognitive, psychopathological, insight, and metacognitive variables as putative predictors of eB2 acceptability.

eB2 App

The eB2 [36] is a "passive" smartphone-based platform [37] available on Android and iPhone operating systems, designed for recording functioning-related data such as mobility (location, distance, speed), physical activity (number of steps), sleep data, and social activity (phone use, active apps, social network data) without the subject's collaboration, that is, running in the background of users' phones (it is passive) other than the initial configuration, which was assisted by one researcher (VGRR). All the above information is gathered from inertial sensors, physical activity, phone calls, message logs, app usage, nearby Bluetooth and Wi-Fi connections, and location. In addition, more detailed activity information and nearby location data can be accessed through Google Play services. Additional resources include a noncontinuous recording schedule and automatic sleep/wake function so that the battery can be safely saved. Further, if eB2 is stopped due to user's actions or failures/reboots, the operating system can relaunch itself. Data are anonymized and sent to a secure server, thereby allowing continuous feedbacks from digital phenotyping [5].

Measures

In terms of sociodemographic variables, we collected data on age, gender, and education level. Regarding clinical variables, diagnosis (schizophrenia-F20 versus all other SSD), previous suicidal behavior (present vs absent), duration of illness (≤5 years vs >5 years), and previous admissions were included. Premorbid adjustment was measured with the premorbid adjustment scale (PAS) [38], which provides 3 scores on childhood, early adolescence, and late adolescence premorbid adjustment (higher scores indicating poorer premorbid adjustment).

XSL•FO
RenderX

Two neurocognitive measures were used, namely, intelligence quotient, which was estimated with the vocabulary subtest of the Wechsler Adults Intelligence Scale-IV [35], and executive function, which was evaluated using the Trail Making Test (TMT) [39]. TMT involves connecting numbers (Task A) or alternating numbers and letters (Task B); therefore, time to complete each task (in seconds) was taken. Subtracting the time to complete task A from time to complete task B provides an overall measure of the executive function (set shifting), having controlled for processing speed (TMT B-A). Psychopathological symptoms were rated using the Spanish version [40] of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia [41].

Clinical insight, which was the primary outcome of the RCT [33], was assessed with the Spanish version [42] of the Schedule for Assessment of Insight [43]; it provides scores on global insight and 3 insight dimensions: illness recognition, symptom relabeling, and treatment compliance. The higher the score, greater the insight. Three metacognitive dimensions were considered, namely, jumping to conclusions (JTC), cognitive insight, and theory of mind (ToM). JTC was measured with the Beads Task [44]. On the basis of probability (in task 1, the probability is 85:15, while in task 2, the probability is 60:40), the individual must decide the jar to which the extracted bead belongs. JTC is considered if a decision is made after extracting 1 or 2 beads. Cognitive insight was measured with the Spanish version [45] of the Beck Cognitive Insight Scale (BCIS) [46], which is a 15-item self-rated scale and yields 2 factors, namely, self-reflectiveness (9 items) and self-certainty (6 items). An overall measure of cognitive insight, that is, composite index, can thus be calculated by subtracting self-certainty from self-reflectiveness. ToM was evaluated by means of the Hinting Task [47] Spanish version [48] by using 2 stories: scores ranged from 0 to 4, where higher scores indicated better ToM performance; and the Emotions Recognition Test Faces activity [49], which is composed of 20 different photographs showing people's emotions and 2 answers at the bottom of each picture, one of each is right and the other is wrong. Scores ranged from 0 to 20; the higher the scores, the better was the ToM performance.

Statistical Analysis

First, we calculated acceptability in terms of (1) recruitment, that is, the proportion of RCT participants who consented to eB2 installation (users), and (2) retention, that is, the proportion of users whose eB2 remained installed till the end of the study period. With regard to the predictors of eB2 acceptability, we first inspected the variable distributions by using histograms and the Kolmogorov–Smirnov test. Second, exploratory bivariate analyses investigated differences between users and nonusers in sociodemographic, premorbid adjustment, clinical, neurocognitive, psychopathological, insight, and metacognitive variables. Parametric and nonparametric tests were used as appropriate. Third, a multivariable binary logistic regression model tested the real influence of the above (independent) variables on "being user" (vs nonuser), that is, acceptability, which was the main outcome measure. A 95% significance level was set for all the above analyses, which were performed using SPSS version 25.0 (IBM Corp).

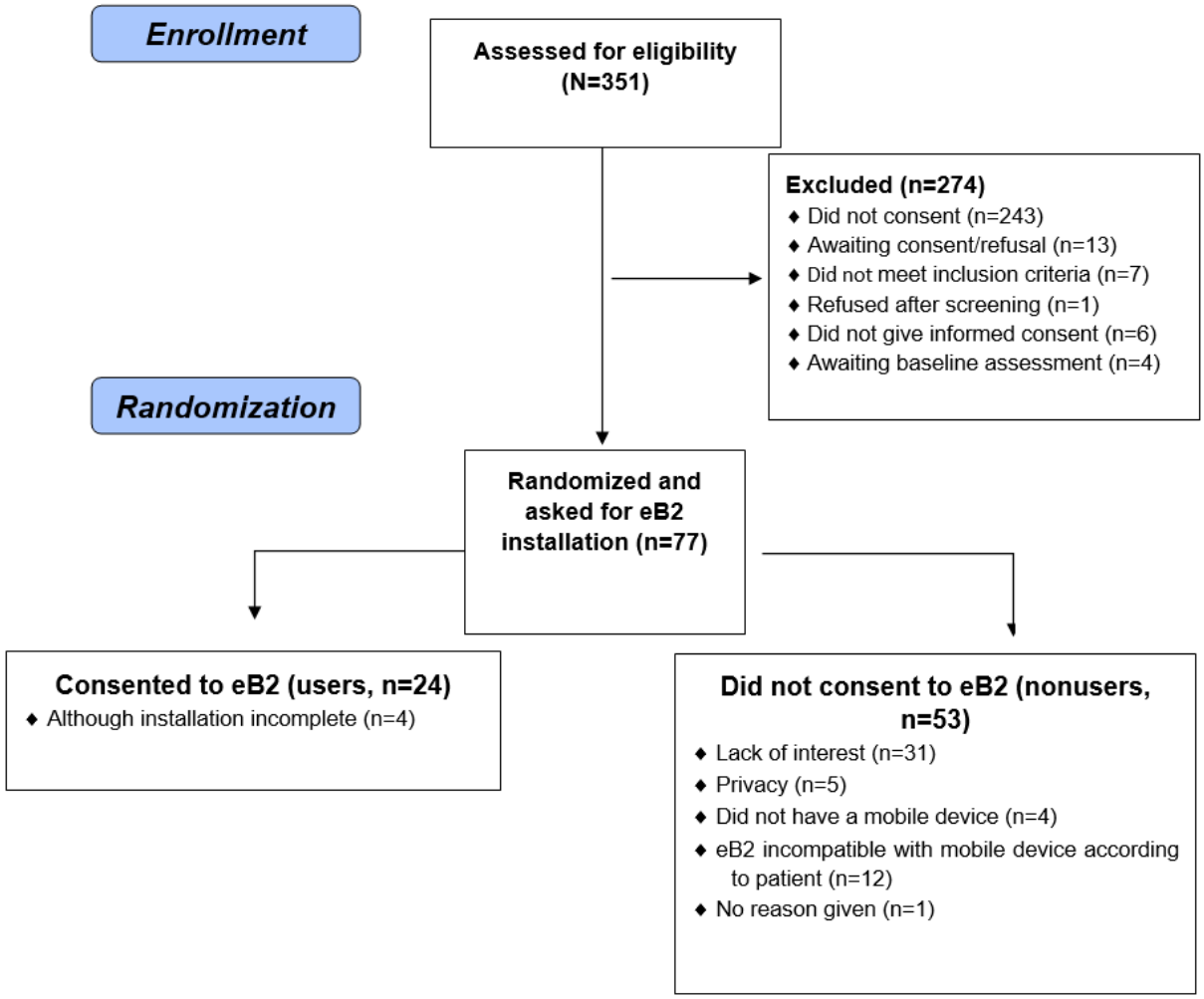
Results

Sociodemographic Characteristics of the Participants

From June 10, 2019 to the March 11, 2020, 351 eligible patients were approached and invited to participate in the RCT by the treating consultant psychiatrist (LMI or SSA) or psychologist (LML). Of these, 77 individuals (22%) agreed to take part in the RCT, met the inclusion criteria, and were available at the baseline assessment; therefore, they were asked to install eB2. Twenty-four subjects (31%) agreed and gave consent (users) (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart in this study. eB2: evidence-based behavior.



Although in 4 users, eB2 installation could not be completed owing to technical issues, in the analyses below, we compared all the individuals who accepted eB2 (users, 24/77, 31%) with those who did not (nonusers, 53/77, 69%). Only 6 subjects uninstalled the app prior to the end of the study, which occurred at a median of 14.50 weeks. Hence, there were 14 users who did not uninstall eB2 (ie, retention was 14/20, 70%) of whom 7 subjects continued to use the app (ie, data being sent to the server) till the end of the study. For the remaining 7 users, there was no evidence of data being sent to the server (eg, they had no wireless access or the battery may have run out without being replaced). Time of eB2 data uploading ranged from 0 to 44 weeks (median 12 weeks).

We asked nonusers (n=53) for the reasons underlying refusal. For descriptive purposes, 4 categories (see below) were created by 2 researchers (JDLM and VGRR) who independently classified participants' free-text verbal responses according to such categories. Any inconsistency was resolved by consensus by 2 researchers (EBG and MLB). Results were (1) lack of interest (31/53, 58%), (2) privacy issues (5/53, 9%), (3) lack of a mobile device (4/53, 7%), and (4) eB2 incompatible with mobile device (prior to installation) according to the patient (12/53, 23%). One individual gave no reason. However, no nonuser raised issues about the configuration/installation of eB2, which was carried out in front of him/her by one researcher (VGRR), that is, at no point was the smartphone taken away from the owner. Time from eB2 installation to last eB2 data registration by the server (n=20) ranged from 0 to 553 days (median 187.5). One subject experienced a psychotic relapse and asked the research team to uninstall the app, which was honored. However, no causality association between eB2 installation and mental state relapse could be established. The sample population characteristics (N=77) and differences between users (n=24) and nonusers (n=53) are presented in Table 1.

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of the participants.

Characteristics	Total sample (N=77)	Users (n=24)	Nonusers (n=53)	*t* (*df*)	χ^2 (*df*)	*P* value
Sociodemographic variables						
Age (years), mean (SD)	47.69 (9.76)	42.33 (10.75)	50.11 (8.30)	–3.46 (75)	N/A[a]	.001
Gender (males), n (%)	41 (53)	14 (58)	27 (51)	N/A	0.4 (1)	.55
Education level (primary), n (%)	13 (17)	1 (8)	12 (23)	N/A	4.0 (1)	.045
Premorbid adjustment [38], mean (SD)						
Childhood	5.80 (3.79)	4.96 (2.76)	6.19 (4.15)	–1.53 (64.49)	N/A	.13
Early adolescence	7.64 (4.64)	5.33 (2.76)	8.71 (4.95)	–3.80 (71.21)	N/A	<.001
Late adolescence	7.69 (4.90)	6.57 (3.07)	8.23 (5.52)	–1.63 (67.33)	N/A	.11
Clinical variables						
Diagnosis, n (%)	48 (62)	16 (67)	32 (60)	N/A	0.3 (1)	.60
Previous suicidal behavior, n (%)	31 (40)	9 (37)	22 (41)	N/A	0.1 (1)	.74
Duration of illness (<5 years), n (%)	8 (10)	2 (8)	6 (11)	N/A	0.2 (1)	.69
Previous admissions, mean (SD)	3.46 (3.99)	2.96 (2.87)	3.68 (4)	–0.72 (74)	N/A	.47
Psychopathology (PANSS)[b] , mean (SD)						
Positive	8.44 (3.67)	9.17 (3.94)	8.11 (3.53)	1.17 (75)	N/A	.25
Negative	14.91 (5.89)	14.00 (5.27)	15.32 (6.16)	0.50 (75)	N/A	.37
Disorganization	6.05 (2.61)	5.50 (2.38)	6.30 (2.69)	–1.25 (75)	N/A	.21
Mania	6.25 (1.86)	6.67 (2.20)	6.06 (1.67)	1.34 (75)	N/A	.18
Depression	6.94 (2.70)	7.29 (2.77)	6.77 (2.67)	0.78 (75)	N/A	.44
Insight (SAI-E)[c] , mean (SD)						
Total insight	15.55 (2.29)	15.67 (5.26)	15.49 (5.35)	0.13 (75)	N/A	.89
Illness recognition	5.36 (2.68)	5.29 (2.77)	5.40 (2.67)	–0.16 (75)	N/A	.87
Symptoms relabeling	5.87 (2.81)	6.21 (2.23)	5.72 (3.05)	0.79 (59.61)	N/A	.43
Treatment compliance	4.31 (1.57)	4.17 (1.58)	4.38 (1.58)	–0.54 (75)	N/A	.59
Neurocognition, mean (SD)						
Intelligence quotient	104.61 (11.72)	105.83 (12.48)	104.06 (11.44)	0.61 (75)	N/A	.54
Trail Making Test A [39]	59.33 (29.50)	45.38 (12.51)	65.90 (30.16)	–4.16 (69)	N/A	<.001
Trail Making Test B	126.25 (56.83)	102.13 (34.27)	138.57 (62.20)	–3.18 (63)	N/A	.002
Trail Making Test B–A	68.91 (43.65)	56.75 (34.68)	75.13 (46.70)	–1.70 (63)	N/A	.09
Metacognition						
Jumping to conclusions (85:15), n (%)	42 (56)	13 (54)	29 (54)	N/A	0.1 (1)	.83
BCIS-SR[d], mean (SD)	15.43 (5.11)	17.18 (3.86)	14.69 (5.41)	1.95 (72)	N/A	.06
BCIS-SC[e], mean (SD)	7.67 (3.42)	7.42 (3.20)	7.80 (3.55)	–0.44 (71)	N/A	.66
BCIS-CI[f], mean (SD)	7.74 (6.66)	9.95 (4.789	6.73 (7.17)	1.92 (68)	N/A	.06
Hinting task, mean (SD)	2.25 (1.33)	2.08 (1.38)	2.32 (1.31)	–0.72 (75)	N/A	.47
ERTF[g], mean (SD)	16.86 (2.16)	16.96 (1.90)	16.81 (2.28)	0.27 (75)	N/A	.79

[a]N/A: not applicable.

[b]PANSS: Positive and Negative Syndrome Scale for schizophrenia [41].

[c]SAI-E: Schedule for Assessment of Insight, expanded version [43].

[d]BCIS-SR: Beck Cognitive Insight Scale-Self-Reflectiveness.

[e]BCIS-SC: Beck Cognitive Insight Scale-Self-Certainty.

XSL•FO
RenderX

[f]BCIS-CI: Beck Cognitive Insight Scale-Cognitive Insight.

[g]ERTF: Emotions Recognition Test Faces [49].

At the time of the study inception, users were younger than nonusers (42.33 [SD 10.75] vs 47.69 [SD 9.76] years, respectively; t_{75}=–3.46; *P*=.001). The proportion of less educated people among users was lower than that among nonusers (1/24, 8% vs 8/53, 23%, respectively; χ^2_1=4.0; *P*=.045). No between-group differences in gender, diagnosis, history of suicidal behavior, illness duration, or number of previous admissions emerged from the analyses. PAS scores among users were lower (ie, better premorbid adjustment) than those among nonusers, which reached significance in early adolescence (5.33 [SD 2.76] vs 8.71 [SD 4.95], respectively; $t_{71.21}$=–3.80; *P*<.001). No significant differences between users and nonusers in psychopathological symptom severity (PANSS factors) or clinical insight scores were found. Although intelligence quotient did not significantly differ between groups, users had better executive function performance, that is, it took them shorter time (in seconds) to complete TMT-A (45.38 [SD 12.51] vs 65.90 [SD 30.16], respectively; t_{63}=–4.16; *P*<.001) and TMT-B (102.13 [SD 34.27] vs 138.57 [SD 62.20], respectively; t_{63}=–3.18; *P*=.002) than nonusers. Differences in time to complete task B minus time to complete task A were nonsignificant (*P*=.09). In terms of metacognitive tasks, the BCIS self-reflectiveness (17.18 [SD 3.86] vs 14.69 [SD 5.41], respectively; t_{72}=1.95; *P*=.06) and the BCIS Composite Index (9.95 [SD 4.78] vs 6.73 [SD 7.17], respectively; t_{68}=1.92; *P*=.06) among users were higher than those among nonusers (indicating better metacognitive performance). Neither JTC (*P*=.83) nor ToM measures (Hinting Task: *P*=.47; Emotions Recognition Test Faces activity: *P*=.79) distinguished users from nonusers significantly.

Binary Multivariable Logistic Regression Model on User (as Outcome)

Age, education level, early adolescence premorbid adjustment, TMT-A, TMT-B, and cognitive insight (only composite index was taken to avoid multicollinearity) were significantly associated with being a user, and they were therefore entered into the binary multivariable logistic regression model (Table 2).

Table 2. Multivariable binary logistic regression model.[a]

Characteristics	Unstandardized coefficient	SE	Wald	*P* value	Odds ratio (95% CI)
Age	–0.075	0.038	3.910	.048	0.928 (0.861-0.999)
Education level	–0.967	1.289	0.563	.45	0.380 (0.030-4.755)
Early premorbid adjustment scale [38]	–0.285	0.110	6.695	.01	0.752 (0.606-0.933)
Trail Making Test A [39]	–0.030	0.025	1.488	.22	0.970 (0.924-1.018)
Trail Making Test B	–0.005	0.010	0.278	.60	0.995 (0.976-1.014)
Cognitive insight	0.062	0.061	1.043	.31	1.064 (0.944-1.200)

[a]Model χ^2_6= 25.3, *P*<.001. The model explained 44.7% (Nagelkerke R^2) of the variance and correctly classified 77% (59/77) of the cases. Specifically, 55% (13/24) of users and 88% (47/53) of nonusers were correctly predicted by the model.

Age (OR 0.93, 95% CI 0.86-0.99; *P*=.048) and early adolescence PAS score (OR 0.75, 95% CI 0.61-0.93; *P*=.01) remained significant. The final model (χ^2_6=25.3; *P*<.001) explained 44.7% (Nagelkerke R^2) of the variance on being a user (or acceptability, ie, the outcome variable) and correctly classified 77% (59/77) of the subjects, that is, 54% (13/24) of the users and 88% (47/53) of the nonusers (Table 2).

Discussion

Main Findings

We used data from an ongoing RCT of metacognitive training with an unselected sample of patients with SSD [33] and we compared those participants who consented to installation (users) of a passive smartphone-based EMA app, eB2, with those who did not (nonusers) in order to investigate eB2 acceptability and the factors that predicted the acceptability. First, as postulated, acceptability was lower than 50% (approximately one-third of the participants), thereby indicating low recruitment. However, contrary to our second hypothesis, retention was higher than expected since most users had not uninstalled eB2 till the end of the follow-up period (at 14 weeks) and in half of the users, there was evidence of continued eB2 use. However, this may have been due to the chronicity of the participants (illness duration was longer than 5 years in almost 90% of them) or negative symptom severity (based on the PANSS ratings). In other words, not having uninstalled the app may well reflect a lack of interest in eB2 rather than the other way round. Our third hypothesis was in part supported by the study results. eB2 users were younger, had a higher education level, better premorbid adjustment and executive function, and higher levels of cognitive insight than nonusers, although only being young and good early adolescence premorbid adjustment survived the multivariable regression model.

Comparison With Previous Literature

Acceptability of eB2 Among Patients With SSD

Previous studies on mobile-based apps in schizophrenia showed good acceptability levels. Back in 2013, a mobile app and text messaging system was tested in 24 subjects: patients with schizophrenia (n=22) and patients with schizoaffective disorder (n=2) (Diagnostic and Statistical Manual of Mental Disorders,

XSL•FO
RenderX

4th Edition criteria), among whom recruitment rate was estimated at approximately 70% [50]. Consistent with this, exploring the mental health app FOCUS use by 33 individuals with schizophrenia was reported to be 61% [51]. However, participants in these studies received a financial incentive, which may raise ethical issues [4], particularly in patients with serious mental illness such as SSD, who tend to have limited incomes [52-54]. Hence, owing to a potential selection bias, caution is needed when interpreting these findings. In keeping with this, acceptability of a smartphone-based app designed for self-reporting psychotic symptoms was significantly lower (50%) when there were no financial incentives [7].

Of note, all the above studies required an active role by participants who had to upload data to the apps themselves. Since patients with SSD tend to have poor compliance [11,55], an alternative to increase EMA acceptability may be using passive smartphone-based apps such as eB2. A passive smartphone EMA tool correlated with in-clinic assessments of sleep quality, including high retention levels (90%) over 6 weeks [23]. Another passive approach to recording functioning data is wearing mHealth devices, which was accepted by the vast majority (80%) of those inpatients with schizophrenia who were found eligible in one study [20]. Consistent with this, 14 out of 15 participants completed an investigation (ie, high retention levels, 93%) on rest/activity recording at sleep time (recruitment was not reported) who received a financial compensation [56]. Regarding patient satisfaction, it is worth noting that 81% of a small sample of patients (n=30) with schizophrenia wearing these devices provided positive or excellent feedback [24].

One may question why we decided to test a passive EMA app in this unselected sample of patients with SSD. As mentioned above, the main aim of this investigation was to test whether metacognitive training can improve insight and clinical outcomes in patients with SSD, that is, addressing noncompliance, which is of major relevance in the psychosis field. In keeping with this, we speculated that a passive EMA app (which required no actions to be taken by the subject) would achieve higher levels of acceptance than an active app. Given the comprehensiveness of the RCT, we considered that asking participants to install an app and upload data (ie, active apps) would make refusal more likely. However, whether passive or active EMA devices may increase or decrease acceptability among patients with psychotic disorders requires further investigation. The extent to which EMA devices may trigger psychotic phenomena in some, but not all, individuals [10] may, in part, explain this. Future studies, free of financial incentives to participants, are therefore warranted to establish what determines this response, including between-individual differences.

Predictors of Acceptability of eB2 Among Patients With SSD

In our sample, users were significantly younger than nonusers. At first glance, one may question the extent to which a between-group difference of 8 years, although statistically significant, is clinically meaningful. In this regard, 2 issues should be noted. First, this finding is in full agreement with that reported in our previous study [2], which found a difference in the age by 6 years between a large group (n=2838) of active users of an active EMA tool, such as MeMind, and nonusers (n=10,973). Second, based on the nQuery method, this difference equates to an effect size of 0.79, that is, a large effect size, which would provide further support for this finding. It is true, however, that the full clinical implications of this result remain subject to further debate. Although it is intuitive to think that youth may be linked with higher levels of eHealth literacy, which may contribute to EMA acceptability, smartphone-based EMA has been successfully implemented in older adults with schizophrenia [57]. Our findings also revealed that users have higher education, better premorbid adjustment, and better executive function performance than nonusers, which is consistent with that reported in a previous study showing a positive association of EMA compliance with neurocognition [15].

Although some previous studies linked noncompliance with negative symptom severity and low premorbid intelligence quotient [17], we failed to replicate this. However, our results were consistent with those of other studies that showed no association of EMA compliance with clinical variables [58] or age, medication, and symptom severity [59]. In line with our results, refusal to participate in one study [18] was associated with clinical appointment nonattendance. Of relevance, while no clinical insight dimension such as illness awareness, symptoms, relabeling, and treatment compliance [60] differed between users and nonusers, cognitive insight was found to predict eB2 acceptability in the bivariate analyses. Cognitive insight [46], which is a core metacognitive domain, may thus become a more relevant predictor of EMA acceptability among subjects with SSD than clinical insight despite the relationship between clinical insight and compliance [11,55]. If the aforementioned association of cognitive insight with EMA acceptability/compliance was replicated, interventions targeting metacognition, such as metacognitive training [61], may increase EMA acceptability/compliance, which remains to be investigated.

Strengths, Limitations, and Implications on Clinical Practice and Future Research

In spite of previous concerns about EMA tool use by patients with SSD, almost one-third of the trial participants agreed to the installation of the eB2 passive mobile app, which opens new directions for clinical practice and research. Participants did not have to take any active role in uploading data or installing the app, which was completed by one researcher (VGRR), which may have increased acceptability levels. EMA-based methods may also pave the way toward remote monitoring of such a vulnerable group of patients. However, further studies should explore more successful strategies aimed at increasing EMA acceptability among patients with SSD. Information leaflets explaining evidence-based benefits from EMA in lay terms within a proper patient-researcher/clinician relationship may contribute to this. Unlike most previous studies [14], participants in this RCT did not receive a financial compensation. Although subject to further debate, by doing so, we may have avoided a potential selection bias. In other words, our findings do reflect the extent to which real-world patients with SSD consented to a mobile-based EMA app, and what appears to underlie this,

XSL•FO
RenderX

regardless of the potential financial incentives. Specifically, the vast majority (31/53, 59%) of those who refused to install eB2 reported lack of interest. Although a financial incentive may have reduced the proportion of nonusers, this would have not reflected a real patient involvement in using EMA.

This study findings should be considered in light of some limitations. First, participants came from an RCT and therefore gave consent and completed a comprehensive set of assessments. Lack of cooperativeness was also an exclusion criterion. Thus, referrers (LLI, SSA, LML) only found 351 patients to be eligible over the study recruitment period, which had to be stopped owing to the COVID-19 outbreak in Spain. Regretfully, we did not systematically record the total number of patients that they saw in clinic, which was much higher. Only 77 agreed to take part in the RCT. Hence, those with poorer insight levels were therefore less likely to take part in this study, which may limit the generalizability of the results. However, this ethical requirement and the subsequent limitation in terms of generalizability applies to most studies on insight in psychosis. Nonetheless, this study is part of an RCT of metacognitive training (detailed above), which was not originally designed to test this study hypotheses. Not only this may have affected the representativeness of the study sample, but also much caution should be taken when applying the study results to patients with SSD in other settings. Second, all participants were mental health service users living in Madrid, which is an inner urban area, and results may not apply to people with psychosis receiving mental health input from primary care (ie, only from the general practitioner) or those residing in rural areas. Third, only a small proportion of the RCT participants agreed to eB2 installation; therefore, some between-group (users vs nonusers) comparisons may have lacked sufficient power. Fourth, other nontested variables may affect eB2 acceptability. Finally, future studies may involve families and carers in EMA app installation and compliance. Regretfully, we did not systematically collect data in the first face-to-face interview with participants in the RCT in terms of variables related to the researchers asking for consent (JDLM and VGRR) and whether their relationship with the participants may have affected eB2 acceptability. This said, we suggest registering information on these variables in future EMA acceptability studies. Further, it should be noted that as per the protocol of the RCT, these EMA users had to have consented to participating in the trial. Hence, we cannot rule out that a number of those who refused to enroll the RCT may have accepted eB2, although this seems unlikely since RCT participants tend to be those individuals with higher levels of cooperativeness.

Our work, therefore, adds to the growing field of e-mental health. Within the context of the COVID-19 pandemic, remote telemedicine-based mental health services need to be prioritized, which is in line with previous mental health policies focused on information and communication technology [62]. Not only the COVID-19 pandemic is likely to have a negative impact on mental health outcomes, including increased suicide rates [63] via unemployment rise [64], but also underfunded/underresourced services will have to continue delivering mental health care to vulnerable patients in need [65]. Although eHealth tools may mitigate this [66], long-term outcomes remain unknown, particularly regarding patients with SSD.

In keeping with this, our results highlight that EMA methods may need to be tailored to patients with SSD. Otherwise, there is a high risk that this group of vulnerable patients may be neglected by newly developed approaches to mental health care with the subsequent very negative impact on outcomes, including increased stigma. Specifically, much attention should be paid to patients with SSD when getting older and to those with low education level, poor premorbid adjustment, and deficits in executive function and cognitive insight. Worryingly, not only are these the most vulnerable individuals affected by such a serious mental illness, but also, based on our results, they appear to be the most reluctant ones to use remote resources such as EMA, which are likely to be prioritized by mental health services in the post-COVID-19 pandemic era. The question, therefore, arises: how can we practically tailor EMA methods to subjects with SSD? First, as noted above, we think that passive apps should be more widely recommended than active apps, given the overall poor compliance in patients with SSD. Second, patients need to be properly reassured that these resources do not control their thoughts or monitor them personally since data are anonymized, which requires a proper doctor-patient relationship. In other words, patients with SSD may be less likely to consent to researchers who are not involved in their clinical care. This said, this is definitely an area in which further research is needed.

More specifically, 2 main interlinked unmet challenges need to be addressed by the so-called e-mental health in the years to come. First, the satisfaction of the patients with SSD toward, and adherence to, new technologies and related devices should be improved. Second, financial incentives should not be considered to achieve this, which may stress further underfunded mental health services, particularly in low-income countries.

Acknowledgments

This study was supported by the Universidad Autónoma de Madrid and European Union via the Intertalentum Project Grant-Marie Skłodowska Curie Actions (GA 713366) to JDLM who is the principal investigator. This grant, therefore, funds both JDLM's salary and the consumable expenses related to the study. MLB, VGRR, ASEM, PJEA, SSA, LMI, LML, and EBG's salaries come from the Hospital Universitario Fundación Jiménez Díaz, where this study is currently being carried out, which therefore provides the necessary institutional/departmental support for its development. Additional departmental support concerning the use of EMA is provided by the Instituto de Salud Carlos III (Madrid, Spain) (ISCIII PI16/01852), the Madrid Regional Government (Madrid, Spain) (B2017/BMD-3740 AGES-CM 2CM; Y2018/TCS-4705 PRACTICO-CM), and the Spanish Ministerio de Ciencia, Innovación y Universidades (TEC2017-92552-EXP). ASD acknowledges funding support from University College

XSL•FO
RenderX

London, which covers his salary. The funders had no role in hypothesis generation, study design, decision to publish, or the manuscript writing. No funder has a conflict of interest in relation to the study's results and findings. The funding bodies supported the peer review for this manuscript.

Authors' Contributions

JDLM is the principal investigator of this RCT, of which this study is part. JDLM, MLB, APS, VGRR, ASEM, PJEA, SSA, LMI, LML, AAR, ASD, and EBG contributed in the process of protocol design, hypothesis generation, manuscript preparation, and fulfilled the criteria for authorship. JDLM wrote the first draft. EBG conceived the study, participated in its design, and implemented the project. MLB, APS, ASD, and EBG contributed to the interpretation of results. All the authors read and approved the final manuscript.

Conflicts of Interest

AAR and EBG cofounded the evidence-based behavior app.

References

1. Measuring digital development: Facts and figures 2020. International Telecommunication Union. URL: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2020-11-12]
2. Barrigón ML, Berrouiguet S, Carballo JJ, Bonal-Giménez C, Fernández-Navarro P, Pfang B, MEmind study group. User profiles of an electronic mental health tool for ecological momentary assessment: MEmind. Int J Methods Psychiatr Res 2017 Mar;26(1):e1554 [FREE Full text] [doi: 10.1002/mpr.1554] [Medline: 28276176]
3. Firth J, Torous J, Yung AR. Ecological momentary assessment and beyond: The rising interest in e-mental health research. J Psychiatr Res 2016 Sep;80:3-4 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.05.002] [Medline: 27236099]
4. Onnela J, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
5. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
6. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 07;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
7. Palmier-Claus JE, Ainsworth J, Machin M, Barrowclough C, Dunn G, Barkus E, et al. The feasibility and validity of ambulatory self-report of psychotic symptoms using a smartphone software application. BMC Psychiatry 2012 Oct 17;12:172 [FREE Full text] [doi: 10.1186/1471-244X-12-172] [Medline: 23075387]
8. Morgan C, Lappin J, Heslin M, Donoghue K, Lomas B, Reininghaus U, et al. Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. Psychol. Med 2014 Feb 26;44(13):2713-2726. [doi: 10.1017/s0033291714000282]
9. Hjorthøj C, Stürup AE, McGrath JJ, Nordentoft M. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 2017 Apr;4(4):295-301. [doi: 10.1016/S2215-0366(17)30078-0] [Medline: 28237639]
10. van der Krieke L, Wunderink L, Emerencia AC, de Jonge P, Sytema S. E-mental health self-management for psychotic disorders: state of the art and future perspectives. Psychiatr Serv 2014 Jan 01;65(1):33-49. [doi: 10.1176/appi.ps.201300050] [Medline: 24129842]
11. Lysaker PH, Pattison ML, Leonhardt BL, Phelps S, Vohs JL. Insight in schizophrenia spectrum disorders: relationship with behavior, mood and perceived quality of life, underlying causes and emerging treatments. World Psychiatry 2018 Feb;17(1):12-23 [FREE Full text] [doi: 10.1002/wps.20508] [Medline: 29352540]
12. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile Phone Ownership and Endorsement of "mHealth" Among People With Psychosis: A Meta-analysis of Cross-sectional Studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
13. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009;18(1):48-57 [FREE Full text] [doi: 10.1002/mpr.276] [Medline: 19195050]
14. Firth J, Torous J. Smartphone Apps for Schizophrenia: A Systematic Review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):e102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
15. Granholm E, Loh C, Swendsen J. Feasibility and validity of computerized ecological momentary assessment in schizophrenia. Schizophr Bull 2008 May;34(3):507-514 [FREE Full text] [doi: 10.1093/schbul/sbm113] [Medline: 17932087]
16. Moitra E, Gaudiano BA, Davis CH, Ben-Zeev D. Feasibility and acceptability of post-hospitalization ecological momentary assessment in patients with psychotic-spectrum disorders. Compr Psychiatry 2017 Apr;74:204-213 [FREE Full text] [doi: 10.1016/j.comppsych.2017.01.018] [Medline: 28231480]

XSL•FO
RenderX

17. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
18. Niendam TA, Tully LM, Iosif A, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: A longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan;96:239-246. [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
19. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological Momentary Assessment and Intervention in the Treatment of Psychotic Disorders: A Systematic Review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
20. Shin S, Yeom C, Shin C, Shin J, Jeong JH, Shin JU, et al. Activity monitoring using a mHealth device and correlations with psychopathology in patients with chronic schizophrenia. Psychiatry Res 2016 Dec 30;246:712-718. [doi: 10.1016/j.psychres.2016.10.059] [Medline: 27836243]
21. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile Phone and Wearable Sensor-Based mHealth Approaches for Psychiatric Disorders and Symptoms: Systematic Review. JMIR Ment Health 2019 Feb 20;6(2):e9819 [FREE Full text] [doi: 10.2196/mental.9819] [Medline: 30785404]
22. So SH, Peters ER, Swendsen J, Garety PA, Kapur S. Detecting improvements in acute psychotic symptoms using experience sampling methodology. Psychiatry Res 2013 Nov 30;210(1):82-88. [doi: 10.1016/j.psychres.2013.05.010] [Medline: 23849758]
23. Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017 Oct 16;3(1):37 [FREE Full text] [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
24. Cella M, Okruszek, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2018 May;195:537-542 [FREE Full text] [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
25. Cho H, Gonzalez R, Lavaysse LM, Pence S, Fulford D, Gard DE. Do people with schizophrenia experience more negative emotion and less positive emotion in their daily lives? A meta-analysis of experience sampling studies. Schizophr Res 2017 May;183:49-55. [doi: 10.1016/j.schres.2016.11.016] [Medline: 27881233]
26. Mote J, Fulford D. Ecological momentary assessment of everyday social experiences of people with schizophrenia: A systematic review. Schizophr Res 2020 Feb;216:56-68. [doi: 10.1016/j.schres.2019.10.021] [Medline: 31874743]
27. Granholm E, Holden JL, Mikhael T, Link PC, Swendsen J, Depp C, et al. What Do People With Schizophrenia Do All Day? Ecological Momentary Assessment of Real-World Functioning in Schizophrenia. Schizophr Bull 2020 Feb 26;46(2):242-251 [FREE Full text] [doi: 10.1093/schbul/sbz070] [Medline: 31504955]
28. Reilly T, Mechelli A, McGuire P, Fusar-Poli P, Uhlhaas PJ. E-Clinical High Risk for Psychosis: Viewpoint on Potential of Digital Innovations for Preventive Psychiatry. JMIR Ment Health 2019 Oct 03;6(10):e14581 [FREE Full text] [doi: 10.2196/14581] [Medline: 31584006]
29. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
30. Depp C, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson T, et al. Mobile Interventions for Severe Mental Illness: Design and Preliminary Data From Three Approaches. The Journal of Nervous and Mental Disease Internet 2010 Oct 2010 Oct 01:715-721. [doi: 10.1097/nmd.0b013e3181f49ea3]
31. Wellness tracker. MeMind. URL: www.memind.net [accessed 2021-07-07]
32. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: A real-world feasibility study. J Affect Disord 2020 Sep 01;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
33. Lopez-Morinigo J, Ruiz-Ruano VG, Martínez ASE, Estévez MLB, Mata-Iturralde L, Muñoz-Lorenzo L, et al. Study protocol of a randomised clinical trial testing whether metacognitive training can improve insight and clinical outcomes in schizophrenia. BMC Psychiatry 2020 Jan 29;20(1):30 [FREE Full text] [doi: 10.1186/s12888-020-2431-x] [Medline: 31996174]
34. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
35. Wechsler D. The psychometric tradition: Developing the wechsler adult intelligence scale. Contemporary Educational Psychology 1981 Apr;6(2):82-85. [doi: 10.1016/0361-476x(81)90035-7]
36. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining Continuous Smartphone Native Sensors Data Capture and Unsupervised Data Mining Techniques for Behavioral Changes Detection: A Case Series of the Evidence-Based Behavior (eB2) Study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e197 [FREE Full text] [doi: 10.2196/mhealth.9472] [Medline: 30530465]
37. Evidence-Based Behavior. URL: http://eb2.tech [accessed 2021-07-07]

XSL•FO
RenderX

38. Cannon-Spoor HE, Potkin SG, Wyatt RJ. Measurement of premorbid adjustment in chronic schizophrenia. Schizophr Bull 1982;8(3):470-484. [doi: 10.1093/schbul/8.3.470] [Medline: 7134891]
39. Reitan RM. Validity Of The Trail Making Test As An Indicator Of Organic Brain Damage. PMS 1958 Dec;8(7):271. [doi: 10.2466/pms.8.7.271-276]
40. Peralta V, Cuesta MJ. Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. Psychiatry Res 1994 Jul;53(1):31-40. [doi: 10.1016/0165-1781(94)90093-0] [Medline: 7991730]
41. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
42. Soriano-Barceló J. Insight assessment in psychosis and psychopathological correlates: validation of the Spanish version of the Schedule for Assessment of Insight - Expanded Version. European Journal of Psychiatry. 2016 Jan 15. URL: https://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S0213-61632016000100004&lng=en&nrm=iso&tlng=en [accessed 2021-07-07]
43. Blackwell B. Treatment Compliance and the Therapeutic Alliance. URL: https://psycnet.apa.org/record/1997-36556-000 [accessed 2021-07-07]
44. Brett-Jones J, Garety P, Hemsley D. Measuring delusional experiences: a method and its application. Br J Clin Psychol 1987 Nov;26(4):257-265. [doi: 10.1111/j.2044-8260.1987.tb01359.x] [Medline: 3427248]
45. Gutiérrez-Zotes JA, Valero J, Cortés MJ, Labad A, Ochoa S, Ahuir M, et al. Spanish adaptation of the Beck Cognitive Insight Scale (BCIS) for schizophrenia. Actas Esp Psiquiatr 2012;40(1):2-9 [FREE Full text] [Medline: 22344490]
46. Beck AT, Baruch E, Balter JM, Steer RA, Warman DM. A new instrument for measuring insight: the Beck Cognitive Insight Scale. Schizophr Res 2004 Jun 01;68(2-3):319-329. [doi: 10.1016/S0920-9964(03)00189-0] [Medline: 15099613]
47. Corcoran R, Mercer G, Frith CD. Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. Schizophr Res 1995 Sep;17(1):5-13. [doi: 10.1016/0920-9964(95)00024-g] [Medline: 8541250]
48. Gil D, Fernández-Modamio M, Bengochea R, Arrieta M. [Adaptation of the Hinting Task theory of the mind test to Spanish]. Rev Psiquiatr Salud Ment 2012;5(2):79-88. [doi: 10.1016/j.rpsm.2011.11.004] [Medline: 22854578]
49. Baron-Cohen S, Wheelwright S, Jolliffe AT. Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. Visual Cognition 1997 Sep;4(3):311-331. [doi: 10.1080/713756761]
50. Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res 2013 Apr 05;15(4):e60 [FREE Full text] [doi: 10.2196/jmir.2328] [Medline: 23563184]
51. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
52. Ashcroft RE. Personal financial incentives in health promotion: where do they fit in an ethic of autonomy? Health Expect 2011 Jun;14(2):191-200 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00664.x] [Medline: 21348904]
53. Groth SW. Honorarium or coercion: use of incentives for participants in clinical research. J N Y State Nurses Assoc 2010;41(1):11-3; quiz 22 [FREE Full text] [Medline: 20882821]
54. Singer E, Couper MP. Do incentives exert undue influence on survey participation? Experimental evidence. J Empir Res Hum Res Ethics 2008 Sep;3(3):49-56 [FREE Full text] [doi: 10.1525/jer.2008.3.3.49] [Medline: 19385770]
55. David AS. Insight and psychosis: the next 30 years. Br J Psychiatry 2020 Sep;217(3):521-523. [doi: 10.1192/bjp.2019.217] [Medline: 31685039]
56. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing Rest-Activity Profiles in Schizophrenia Using Wearable and Mobile Technologies: Development, Implementation, Feasibility, and Acceptability of a Remote Monitoring Platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
57. Kimhy D, Myin-Germeys I, Palmier-Claus J, Swendsen J. Mobile assessment guide for research in schizophrenia and severe mental disorders. Schizophr Bull 2012 May;38(3):386-395 [FREE Full text] [doi: 10.1093/schbul/sbr186] [Medline: 22287280]
58. Kumar D, Tully LM, Iosif A, Zakskorn LN, Nye KE, Zia A, et al. A Mobile Health Platform for Clinical Monitoring in Early Psychosis: Implementation in Community-Based Outpatient Early Psychosis Care. JMIR Ment Health 2018 Feb 27;5(1):e15 [FREE Full text] [doi: 10.2196/mental.8551] [Medline: 29487044]
59. Edwards CJ, Cella M, Emsley R, Tarrier N, Wykes THM. Exploring the relationship between the anticipation and experience of pleasure in people with schizophrenia: An experience sampling study. Schizophr Res 2018 Dec;202:72-79 [FREE Full text] [doi: 10.1016/j.schres.2018.06.040] [Medline: 30007868]
60. David AS. Insight and psychosis. Br J Psychiatry 1990 Jun;156:798-808. [doi: 10.1192/bjp.156.6.798] [Medline: 2207510]
61. Moritz S, Woodward T. Metacognitive training for schizophrenia patients (MCT): A pilot study on feasibility, treatment adherence, and subjective efficacy. German Journal of Psychiatry. 2007 Aug 01. URL: https://psycnet.apa.org/record/2008-12205-003 [accessed 2021-07-07]

XSL•FO
RenderX

62. Arango C. Lessons Learned From the Coronavirus Health Crisis in Madrid, Spain: How COVID-19 Has Changed Our Lives in the Last 2 Weeks. Biol Psychiatry 2020 Oct 01;88(7):e33-e34 [FREE Full text] [doi: 10.1016/j.biopsych.2020.04.003] [Medline: 32381276]
63. Kawohl W, Nordt C. COVID-19, unemployment, and suicide. Lancet Psychiatry 2020 May;7(5):389-390 [FREE Full text] [doi: 10.1016/S2215-0366(20)30141-3] [Medline: 32353269]
64. Oyesanya M, Lopez-Morinigo J, Dutta R. Systematic review of suicide in economic recession. World J Psychiatry 2015 Jun 22;5(2):243-254 [FREE Full text] [doi: 10.5498/wjp.v5.i2.243] [Medline: 26110126]
65. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
66. Torous J, Wykes T. Opportunities From the Coronavirus Disease 2019 Pandemic for Transforming Psychiatric Care With Telehealth. JAMA Psychiatry 2020 Dec 01;77(12):1205-1206. [doi: 10.1001/jamapsychiatry.2020.1640] [Medline: 32391857]

Abbreviations

BCIS: Beck Cognitive Insight Scale
eB2: evidence-based behavior
EMA: ecological momentary assessment
JTC: jumping to conclusions
PANSS: Positive And Negative Syndrome Scale for schizophrenia
PAS: premorbid adjustment scale
RCT: randomized controlled trial
SSD: schizophrenia spectrum disorders
TMT: trail making test
ToM: theory of mind

Edited by R Kukafka; submitted 16.12.20; peer-reviewed by J Mote, P Harvey, M Bulgheroni; comments to author 14.01.21; revised version received 10.02.21; accepted 13.05.21; published 26.07.21.

Please cite as:
Lopez-Morinigo JD, Barrigón ML, Porras-Segovia A, Ruiz-Ruano VG, Escribano Martínez AS, Escobedo-Aedo PJ, Sánchez Alonso S, Mata Iturralde L, Muñoz Lorenzo L, Artés-Rodríguez A, David AS, Baca-García E
Use of Ecological Momentary Assessment Through a Passive Smartphone-Based App (eB2) by Patients With Schizophrenia: Acceptability Study
J Med Internet Res 2021;23(7):e26548
URL: https://www.jmir.org/2021/7/e26548
doi:10.2196/26548
PMID:34309576

©Javier-David Lopez-Morinigo, María Luisa Barrigón, Alejandro Porras-Segovia, Verónica González Ruiz-Ruano, Adela Sánchez Escribano Martínez, Paula Jhoana Escobedo-Aedo, Sergio Sánchez Alonso, Laura Mata Iturralde, Laura Muñoz Lorenzo, Antonio Artés-Rodríguez, Anthony S David, Enrique Baca-García. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

App-Tailoring Requirements to Increase Stress Management Competencies Within Families: Cross-sectional Survey Study

Laura Luise Bischoff[1], MA; Hannes Baumann[1], MEd; Charlotte Meixner[1], MA; Patricia Nixon[2], MSc; Bettina Wollesen[1,3], Prof Dr

[1]Department of Movement Science, Faculty of Psychology and Human Movement, University of Hamburg, Hamburg, Germany

[2]Fitbase Institute for Online Prevention GmbH, Hamburg, Germany

[3]Biopsychology and Neuroergonomics, Technical University of Berlin, Berlin, Germany

Corresponding Author:
Laura Luise Bischoff, MA
Department of Movement Science
Faculty of Psychology and Human Movement
University of Hamburg
Mollerstraße 10
Hamburg, 20148
Germany
Phone: 49 40 42838 93
Email: laura.bischoff@uni-hamburg.de

Abstract

Background: Families experiencing high levels of psychological distress are considered a particularly vulnerable population for adverse effects on mental and physical health. Moreover, highly stressed individuals engage less in mental health promoting activities and show low stress management competencies. App-based stress interventions seem promising for the treatment and prevention of stress outcomes and might be a low-threshold solution.

Objective: The aim of this study was to identify the requirements for a tailored app to reduce stress in a cohort of highly stressed families that have low stress management skills.

Methods: Parents (n=1008; age: mean 47.7 years, SD 6.1; female: 599/1008, 59.7%) completed an extensive web-based survey and were subdivided into a target (stressed individuals with low stress competency) and nontarget group according to their reported stress level and stress management competencies. Group differences were analyzed using analysis of variance. In principal component analysis with Kaiser varimax rotation, personally defined stress management goals were grouped into components. Linear regression models were also calculated.

Results: A 3-factor solution cumulatively explained 56% of the variance in personally defined goals of interest for stress management with (1) active strategies (25.61% explained variance), (2) general competency (17.95% explained variance) and (3) passive strategies (12.45% explained variance). The groups differed in age ($F_{1,978}$=27.67, P<.001), health index ($F_{1,958}$=246.14, P<.001), personally defined general-competency goal ($F_{1,958}$=94.16 P<.001), as well as “information acquisition” ($F_{1,971}$=14.75, P<.001) and “need for stimulation” ($F_{1,981}$=54.49, P<.001) personality traits. A regression model showed that for the active strategies goals of interest, only app feature information or instructional videos had a significant effect (P=.02). The general competency factor showed none, and the passive strategies factor showed significant effects for 2 app features—suggestions for planning possible activities with the family (P=.01) and diaries for documentation and development of strategies (P=.03).

Conclusions: The results of this survey study highlight the need to develop an app to increase stress management competencies that takes into consideration perceived stress level, stress management skills, personality, and personally defined goals of the user. The content of the app should be tailored to previously detected personality traits, especially selective information acquisition and low need for stimulation. Furthermore, personally defined stress management goals seem to affect interest in some features.

(J Med Internet Res 2021;23(7):e26376) doi:10.2196/26376

KEYWORDS

mhealth; ehealth; mobile applications; stress management; app features; gamification; family; personality traits

XSL•FO
RenderX

Introduction

Background

Stress is associated with heightened risks of adverse physical and mental health consequences, such as impaired sleep [1], gastrointestinal diseases [2], diabetes [3], coronary heart disease [4], or depression [5]. These consequences are a tremendous burden from a societal, personal, and economical perspective. Families experiencing high levels of psychological distress are considered a vulnerable population [6]. Melchior et al [7] found, for example, that participants who are simultaneously exposed to elevated levels of work stress and high family demands have heightened rates of sickness absence due to psychiatric disorders. Studies investigating work–family spillover effects show that perceived stress at work can be transferred to family members [8-10]. In accordance with the work–family spillover theory, parents play a significant role in their children's health and coping by implementing or reinforcing certain behaviors [11,12]. Family stress was, for example, predictive of less adequate child dietary intake, with one effect occurring indirectly via impaired parent–child relationship quality [13]. In general, health is subject to sociostructural and milieu-specific dependencies for which the family is an important influencing factor [14-18]. It, therefore, seems to be of the utmost importance to create effective interventions to manage high stress levels in families.

Stress Management Competencies

An increasing amount of literature suggests that interventions using different stress management techniques, such as mindfulness, lead to significant psychological health benefits in a wide range of populations [19-22]. Various stress management techniques have been applied and evaluated in diverse populations over the last decades in in-person settings or digital interventions. Active techniques, such as physical activity, can lead to a reduced perception of stress. Regular endurance and strength training, as well as yoga, have been shown to be effective in reducing stress [23-25] as well as acting as buffer against stress appraisal in times of elevated stress [25-27]. Similar results can be found for breathing exercises [28,29] and mindfulness training [30,31], with heterogeneous results for meditation exercises [32,33]. However, these active stress management techniques require regular practice. In comparison, passive but effective ways of managing stress and improving well-being are wellness and sauna [34] and spending time in nature [35,36].

In this study, we define the subjective ability to apply and perform such stress management techniques according to personal demands and stress level as *stress management competencies*. It should, however, be noted that highly stressed individuals are less likely to engage in mental health promoting activities [37-40]. Consequently, families reporting high stress levels presumably also have less stress management competencies. Thus, low-threshold options are needed (1) to support family members experiencing high levels of stress and (2) to teach stress management techniques.

Tailoring in Mobile Health Stress Reduction Interventions

In an increasingly computer-educated European population, information and communication technology might provide unique and low-threshold opportunities to engage parents and families in mobile health (mHealth)–related services and encourage behavior change, to improve health and reduce stress [41-43].

New concepts, such as the PSYCHE system [44], have emerged as technology aids in order to improve or sustain mental health or stress monitoring [45]. Such wearables include personal health records and are designed to encourage health-related behaviors. Various mobile interventions with different guidance formats (eg, self-help, adherence-focused, eCoaching) have been developed to date and have been shown to be effective in the treatment of diabetes [46], depression [42,47], or sleep disorders [48]. Moreover, web-based stress management interventions seem promising for the treatment and prevention of detrimental stress-related outcomes [49]. Nevertheless, 2 meta-analyses [42,49] show that apps incorporating cognitive behavioral therapy or aspects of mindfulness training yield heterogeneous results. In fact, one of the biggest concerns about the usage of mobile interventions for health promotion is low adherence, which can be associated with reduced effectiveness [50,51]. For this reason, research has called for the examination of suitable components that could help to overcome this challenge.

Tailoring [52] was identified as having positive effects on the health outcomes of web-based interventions. Tailoring is defined as

> *any combination of information or change strategies intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and have been derived from an individual assessment. [53]*

A meta-analysis on tailored print health behavior change interventions has demonstrated that tailored messages were superior compared to generic messages and were associated with larger effect sizes [52]. Moreover, variables such as gender or ethnicity did not moderate this effect which underlines the potential of tailored health communication to raise health-related awareness and knowledge about health for various target populations. To further capture the impact of tailoring, research has expanded to using the web as delivery mode, which again demonstrated the superiority against nontailored interventions [54,55]. Next to personalized messages, tailored web-based interventions often include gamification elements such as receiving rewards or social comparisons [56]. A comprehensive systematic review identified engagement promotion and enhancement of effectiveness as main reasons for the application of gamification [57]. Another systematic review on gamification demonstrated that, on average, only one gamification element, such as stories, themes, or display of progress was applied in web-based mental health interventions, with a maximum of 3 applied [56]. Altogether, these studies underline the vast opportunities for tailoring and the inclusion of gamification features and that users might perceive such interventions as

XSL•FO
RenderX

more personally relevant and credible which again could have a significant impact on health outcomes.

Research on tailored web-based stress management interventions is scarce, yet tailoring could be an effective tool to empower users in their stress management skills and to reinforce self-determined health-related behaviors.

Personality Traits

On the other hand, studies show that certain personality traits are associated with specific coping behaviors [58,59] and app usage behaviors [60,61], as well as the response to gamification elements [62]. Individuals with personality factors such as high neuroticism, for example, show more vulnerability toward high stress values and problem coping strategies such as wishful thinking, withdrawal, and emotion-focused coping [58,59]. Notably, personality might predict coping strategies in highly stressed samples more accurately than in less stressed samples [58]. Furthermore, conventional personality theories such as the Big Five Personality theory focus on cognitive or emotional contents to explain motivation and self-regulation. The Personality System Interaction (PSI) theory, on the other hand, focuses on:

> *functional relationships among affective and cognitive macrosystems, i.e., the dynamic processes that underlie human mental functioning [63]*

and might be more suitable to detect and predict self-regulation and volitional aspects of health behavior. PSI theory distinguishes between 2 emotional—(1) the need for stimulation and (2) the need for security—and 2 cognitive systems—(1) the need for information and (2) information processing [64]. Meixner et al [65] investigated the associations between personality traits assessed via PSI theory, interest in app-based monitored physical activity goals, app features, and gamification in order to create tailored mHealth content and found no significant interaction. Furthermore, they concluded that the problem of inactive participants should, in fact, be addressed with app features and gamification elements in accordance with their prior defined goals rather than with their personality traits. Nevertheless, with respect to earlier studies suggesting personality traits being associated with higher stress vulnerability and potentially different stress management competencies, we hypothesize that the results by Meixner et al [65] might not be transferable to stress outcomes and tailored app features.

Study Objectives

To develop and implement a tailored mobile app that appropriately reaches vulnerable und highly stressed families in order to improve their stress management competencies, the following aspects should be addressed: (1) existing stress management techniques and perceived stress management competencies in families, (2) the influence of personality traits, and (3) potentially suitable features for a mobile stress management intervention such as gamification and (4) defined goals of interest in order to individually manage stress.

Previous studies have focused on evaluation of the usage and tailoring effectiveness; however, evidence on the assessment of users' needs and preferences is limited. Given the adverse impact of low adherence on treatment outcomes, understanding technological and content-related factors is crucial for the design and large-scale implementation of app-based stress reduction interventions into routine health care, and ultimately, to help users to interact in a health-promoting way. With respect to the health impairing consequences of high stress levels for each family member, it seems highly relevant to evaluate the families' needs and preferences for mHealth approaches.

Therefore, the aim of this exploratory study was to identify the requirements of an individualized app to reduce stress in a cohort of highly stressed families that have low stress management skills.

The main research questions were (1) Which characteristics can be identified that describe stressed individuals with low stress competency? (2) Which app features and gamification elements are of the most interest for highly stressed participants with low stress-management competency? (3) Which app features and gamification elements are relevant for different types of stress management goals?

Methods

Study Design

This cross-sectional study was part of a project that aims to develop a tailored mHealth intervention for family members and to design health promotion in a sustainable manner. This study was approved by the University of Hamburg ethics committee (file reference: AZ: 2019_270).

Sample

Every family insured by a small German health insurance cooperative (approximately 18,000 families) was invited by post to participate in a web-based survey. Participation in the study was voluntary, in accordance with the principles for medical research involving humans, and participation was not rewarded in any way. The questionnaire development process included several team-internal evaluation procedures and was implemented using Questback software [66], which allowed individual access via QR code. To avoid bias due to involuntary disclosure of sensitive information, there was a no disclosure option for each question. There were no mandatory questions for data protection reasons.

Measures

The questionnaire is available as Multimedia Appendix 1.

Sociodemographic and Health Variables (8 items)

Age in years (1 item) and gender (1 item) were assessed. In order to provide a holistic framework, the concept *health behavior* was based on self-assessment in the dimensions of physical activity (2 items), dietary behavior (2 items), and stress (2 items).

For dietary behavior and physical activity, in each case, 2 questions from the CALO-RE taxonomy of behavior change [67] and the Baecke questionnaire [68] for measuring habitual physical activity were used in combination with the reference values of the World Health Organization [69] and the German Society for Nutrition [70].

XSL•FO
RenderX

A health behavior index was developed based on the physical activity, dietary behavior, and stress questions. For this purpose, each question was first evaluated on a scale of not achieved (0), partially achieved (1) and achieved (2) based on the reference values mentioned above. These values were added, resulting in a score between 0 and 12—if all questions were consistent with the proposed reference values in all 3 dimensions, a person reached an overall health index of 12. This means that the higher the health index, the more health-promoting a person's behavior.

Personally Defined Goals of Interest for Stress Management (10 items)

The following items were extracted from qualitative interviews: performance of meditation exercises, performance of breathing exercises, performance of yoga exercises, performance of mindfulness exercises, performance of relaxation exercises, improvement of stress management competencies, improvement of the ability to perform stress management techniques from anywhere, improvement of personal resilience to stress, spending time in nature, benefit from wellness and sauna offers. We first conducted interviews and then developed a quantitative survey with items extracted from the interviews. A query of the interest for these items was conducted using multiple checkboxes.

Personality Variables (16 items)

The personality questions were derived from previous qualitative interviews and checked for construct validity using the Visual Questionnaire [65]. The personality analysis included health-specific questions, which resulted in a manifestation of 4 personality dimensions (need for security, information acquisition, need for stimulation, and information processing). Each dimension was described by 4 items, each rated on a 6-point Likert scale that ranged from 1 (agree) to 6 (disagree).

App Feature Variables (9 items)

These variables were integrated into the web-based survey to identify tailoring requirements in accordance with our exploratory approach. The questions focused on preferences, ideas, and needs of the respondents in order to design an app in a user-friendly way that was adapted to their needs. The items were individualization of app content, fulfilling common weekly goals and tasks, connecting the app with wearables, increasing knowledge about a healthy lifestyle, suggestions for activities with the family, diaries for documentation and development of strategies, reminders of goals, informational or instructional videos, and analog format for children. Each item was rated by participants on a 6-point scale that ranged from 1 (totally irrelevant) to 6 (totally relevant).

Gamification Feature Variables (14 items)

Questions asking which gamification elements respondents found appealing—comparison with others, in a ranking or on a high score list; controlling and checking progress; collecting points for performance; collecting shared points with other family members; receiving awards, recognition, or encouragement; monetary incentives for achieving goals; linking to the bonus program of the health insurance company; designing an avatar; completing tasks under time pressure, for example, a countdown; advancing to another level or increasing the level of difficulty; sharing and comparing my achieved goals with others; an accompanying storyline; receiving auditory, haptic, or visual feedback; and rating other family members—were rated on a 6-point scale from 1 (would not appeal to me) to 6 (would appeal to me very much). Similarly, these were integrated into the web-based survey in an exploratory manner.

Procedures

The web-based questionnaire (EFS Questback; 2019 version [66]) was preceded by participant information including instructions on anonymity, voluntariness, and data privacy. The participants received an invitation by post to complete the questionnaire. Completing the questionnaire took approximately 30 minutes. Only fully completed surveys were included in analysis.

Statistical Analysis

We used SPSS software (version 27.0; IBM Corp) for statistical analyses.

Step 1

All variables of the questionnaire that asked for personally defined goals of interest for stress management were factor-analytically reduced to 3 factors (active strategies, general competency, passive strategies) in principal component analysis with Kaiser varimax rotation. Bartlett and Kaiser-Meyer-Olkin measure of sampling adequacy tests were performed to test the suitability of variables for factor analysis.

Step 2

Perceived stress level and stress-management competency variables were dichotomized in order to identify stressed individuals with low stress competency as a target group. The characteristics of the target group and the rest of the participants were descriptively characterized. We compared groups using analysis of variance.

Step 3

In order to analyze which app and gamification characteristics are relevant for stressed individuals with low stress competency, the data set was then reduced to only those participants, and to reduce data to relevant variables, all feature and gamification variables with a mean value <3.5 in the target group, indicating irrelevant features, or that did not differ significantly between groups were excluded.

Step 4

We performed 2-way correlation analysis between app and gamification feature variables not excluded in step 3 and the 3 stress reduction target factors (active strategies, general competencies, passive strategies) from step 1.

Step 5

Three linear regression models were calculated, each with 1 of the 3 target factors for stress management strategies (active strategies, general competencies, passive strategies) as a dependent variable. As independent variables, the remaining variables from step 3 (individualization of app content, fulfilling common weekly goals and tasks, increasing knowledge about a healthy lifestyle, suggestions for activities with the family,

XSL•FO
RenderX

diaries for documentation and development of strategies, reminders for objectives, informational or instructional videos, and controlling and checking progress) were included.

Power

In order to be able to demonstrate the anticipated small effect sizes (<0.05) in a multiple linear regression model with 95% power and 8 predictors, a minimum sample size of 463 was calculated (G*Power; version 3.1 [71]).

Data Exclusion

Only fully completed questionnaires were included. For bi- and multivariate analysis procedures, list-wise case exclusion was used.

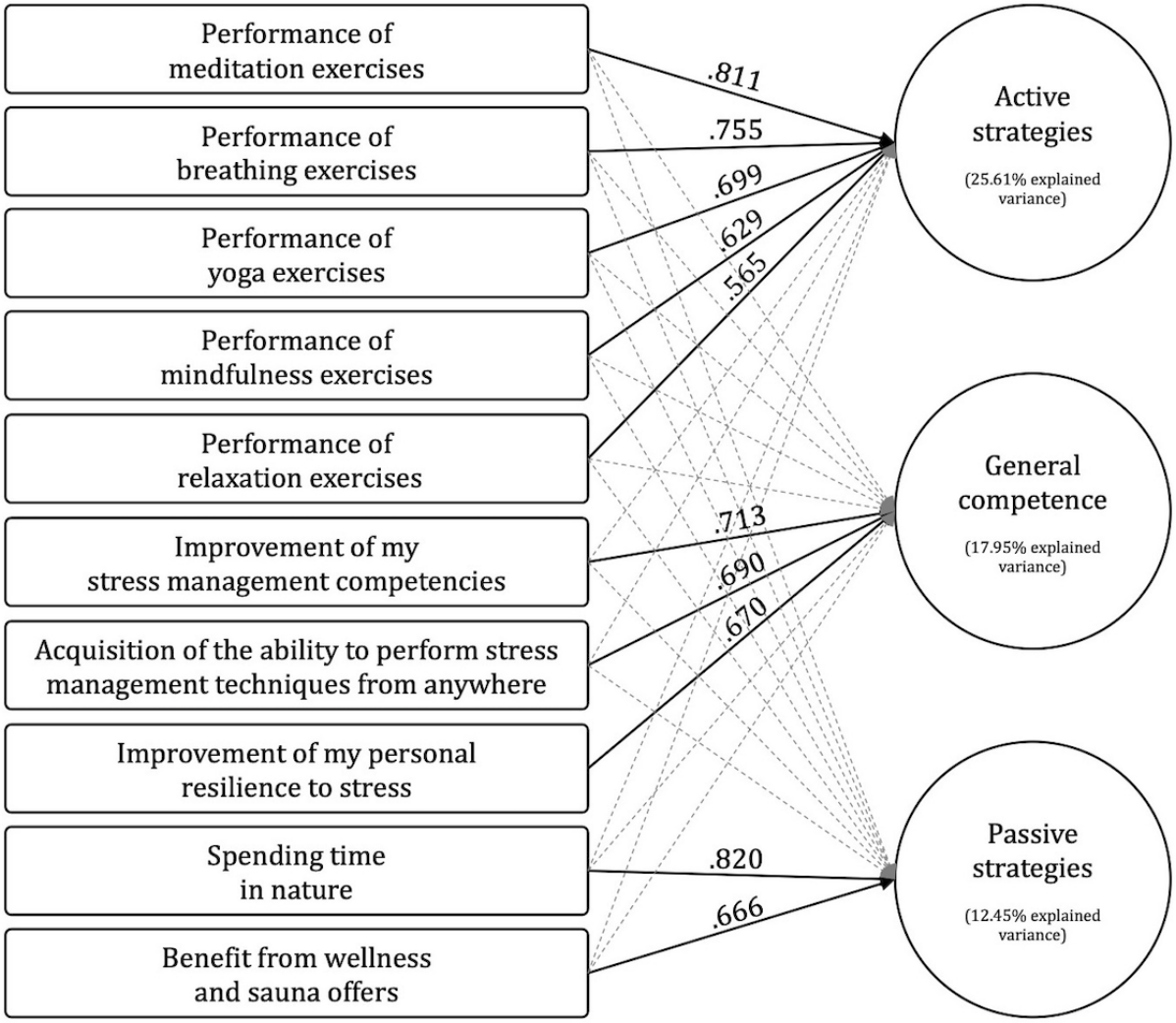
Results

Of 18,000 families invited by post to participate in the web-based survey, 1008 families completed the questionnaire (a response rate of 17.86%). The total sample consisted of 599 female, 398 male, and 7 diverse participants; 4 participants did not give any gender information. The average age of respondents was 47.79 years (SD 6.13).

Factor-Analytical Reduction of Personally Defined Goals of Interest for Stress Management

Both the Bartlett test (χ^2_{45}=2105.563, P<.001) and measure of sampling adequacy (Kaiser-Meyer-Olkin .854) revealed that 10 stress-related target variables (Figure 1) were suitable for factor analysis. Principal component analysis, with varimax rotation indicated the presence of 2 factors with eigenvalues greater than 1.0, and a 3-factor solution that cumulatively explained 56% of the variance was chosen based on the scree plot (and theoretical considerations).

Figure 1. Rotated component matrix of the 10 stress-related target variables. Significant correlations are indicated by a continuous line.



Characteristics

A comparison of participants with low perceived stress, high stress management skills, or both versus participants with high perceived stress and low stress management (Table 1) demonstrated groups differed in age ($F_{1,978}$=21.67, P<.001, ηp^2=.022), health index ($F_{1,958}$=246.14, P<.001, ηp^2=.214),

XSL•FO
RenderX

active strategies ($F_{1,958}$=8.03, P=.01, ηp^2=.008), general competency ($F_{1,958}$=94.16 P<.001, ηp^2=.086), information acquisition ($F_{1,971}$=14.75, P<.001, ηp^2=.053), and need for stimulation ($F_{1,981}$=54.49, P<.001, ηp^2=.012). App feature and gamification variables that met Step 3 criteria were individualization of app content ($F_{1,977}$=8.95, P<.001, ηp^2=.009), fulfilling common weekly goals and tasks ($F_{1,994}$=7.80, P=.01, ηp^2=.008), increasing knowledge about a healthy lifestyle ($F_{1,991}$=9.06, P<.001, ηp^2=.009), suggestions for activities with the family ($F_{1,993}$=10.52, P<.001, ηp^2=.010), diaries for documentation and development of strategies ($F_{1,990}$=12.43, P<.001, ηp^2=.012), reminders for objectives ($F_{1,995}$=4.55, P=.03, ηp^2=.005), informational or instructional videos" ($F_{1,994}$=4.71, P=.03, ηp^2=.005), and controlling and checking progress ($F_{1,998}$=6.82, P=.01, ηp^2=.007).

XSL•FO
RenderX

Table 1. Comparison of sociodemographic, personality, app feature, and gamification feature variables between groups.

Variables	Nontarget group (n=548)		Target group (n=460)		Comparison		
	n	Mean (SD)	n	Mean (SD)	*F* test *(df1,df2)*	*P* value	ηp^2
Sociodemographic variables							
Age (years)	534	48.62 (6.06)	446	46.81 (6.09)	21.67 (1,978)	<.001	.022
Gender						__ [a]	—
Female	294	—	305	—	—		
Male	250	—	148	—	—		
Diverse	3	—	4	—	—		
Health behavior index	525	6.64 (1.71)	435	5.06 (1.36)	246.14 (1,978)	<.001	.214
Personally defined goals for stress management							
Active strategies	548	−.08 (0.04)	460	.09 (1.01)	8.03 (1,978)	.01	.008
General competency	548	−.26 (0.04)	460	.32 (0.87)	94.16 (1,978)	<.001	.086
Passive strategies	548	−.02 (0.04)	460	.03 (1.01)	0.537 (1,978)	.46	.001
Personality variables							
Need for security	526	3.94 (0.83)	448	3.87 (0.83)	2.18 (1,972)	.14	.002
Information acquisition	525	3.66 (0.57)	448	3.52 (0.54)	14.75 (1,971)	<.001	.015
Need for stimulation	531	3.55 (0.62)	452	3.26 (0.61)	54.49 (1,971)	<.001	.053
Information processing	526	4.10 (0.59)	451	4.07 (0.58)	0.68 (1,975)	.41	.001
App feature variables							
Individualization of app content	543	4.27 (1.76)	456	4.59 (1.57)	8.95 (1,978)	<.001	.009
Fulfilling common weekly goals and tasks	541	3.78 (1.58)	455	4.05 (1.48)	7.8 (1,978)	.01	.008
Connecting the app with wearables	534	3.02 (1.74)	448	3.38 (1.81)	10.07 (1,978)	<.001	.010
Increasing knowledge about a healthy lifestyle	538	4.09 (1.62)	455	4.38 (1.44)	9.06 (1,978)	<.001	.009
Suggestions for activities with the family	541	3.61 (1.61)	454	3.94 (1.53)	10.52 (1,978)	<.001	.010
Diaries for documentation and development of strategies	540	3.45 (1.57)	452	3.80 (1.51)	12.43 (1,978)	<.001	.012
Reminders for objectives	542	4.04 (1.58)	455	4.24 (1.47)	4.55 (1,978)	.03	.005
Informational or instructional videos	540	4.08 (1.60)	456	4.30 (1.47)	4.71 (1,978)	.03	.005
Analog format for children	548	1.72 (0.45)	460	1.68 (0.47)	2.56 (1,978)	.11	.003
Gamification feature variables							
Comparison with others, in a ranking or on a high score list	544	2.26 (1.61)	454	2.26 (1.65)	0 (1,976)	.98	0
Controlling and checking my progress	546	4.25 (1.76)	454	4.53 (1.60)	6.82 (1,998)	.01	.007
Collecting points for my performance	545	3.60 (1.83)	455	3.79 (1.78)	2.66 (1,998)	.10	.003
Collecting shared points with other family members	541	3.50 (1.83)	453	3.59 (1.87)	0.61 (1,992)	.43	.001
Receiving awards, recognition, or encouragement	541	3.26 (1.77)	455	3.42 (1.76)	2.21 (1,994)	.14	.002
Providing monetary incentives for achieving goals	536	3.52 (1.88)	453	3.74 (1.81)	3.52 (1,987)	.06	.004
Linking to the bonus program of the health insurance company	541	4.16 (1.93)	456	4.40 (1.80)	3.97 (1,995)	.05	.004
Designing an avatar	541	2.46 (1.64)	454	2.72 (1.77)	5.69 (1,993)	.02	.006

XSL•FO
RenderX

Variables	Nontarget group (n=548)		Target group (n=460)		Comparison		
	n	Mean (SD)	n	Mean (SD)	*F* test *(df1,df2)*	*P* value	ηp^2
Completing tasks under time pressure (eg, a countdown)	543	2.29 (1.50)	456	2.41 (1.58)	1.59 (1,997)	.21	.002
Advancing to another level or increasing the level of difficulty	542	3.22 (1.75)	457	3.39 (1.71)	2.38 (1,997)	.12	.002
Sharing and comparing my achieved goals with others	542	2.18 (1.47)	456	2.22 (1.50)	0.22 (1,996)	.64	0
An accompanying storyline	541	2.65 (1.70)	454	2.73 (1.64)	0.52 (1,993)	.47	.001
Receiving auditory, haptic, or visual feedback	539	3.06 (1.73)	455	3.27 (1.69)	3.9 (1,992)	.05	.004
Rating other family members	539	2.43 (1.60)	454	2.46 (1.62)	0.05 (1,991)	.82	0

[a]Data not provided.

Correlations Between Personally Defined Goals of Interest for Stress Management and App Features in the Target Group

The personally defined *active strategies* factor was correlated with 5 of the 8 features (increasing knowledge about a healthy lifestyle, suggestions for activities with the family, diaries for documentation and development of strategies, reminders for objectives, and informational or instructional videos) (Table 2). The *general competency* factor was correlated with fulfilling common weekly goals and tasks, diaries for documentation and development of strategies, and reminders for objectives. The *passive strategies* factor showed the lowest correlations with the features; it was only correlated with suggestions for activities with the family and diaries for documentation and development of strategies. While some features were correlated with several target factors, others were specific to one factor.

Table 2. Correlations between personally defined goals and app features in stressed individuals with low stress competency.

Feature variables[a]	Active strategies			General competency			Passive strategies		
	r	*P* value	n	*r*	*P* value	n	*r*	*P* value	n
Individualization of app content	0.044	.35	456	0.037	.43	456	0.076	.11	456
Fulfilling common weekly goals and tasks	0.086	.07	455	0.107	.02	455	0.068	.15	455
Increasing knowledge about a healthy lifestyle	0.143	.002	455	0.066	.16	455	0.081	.08	455
Suggestions for activities with the family	0.158	.001	454	0.084	.07	454	0.152	.001	454
Diaries for documentation and development of strategies	0.136	.004	452	0.104	.03	452	0.125	.008	452
Reminders for objectives	0.145	.002	455	0.135	.004	455	0.048	.30	455
Informational or instructional videos	0.201	<.001	456	0.090	.05	456	0.061	.19	456
Controlling and checking progress	0.030	.52	454	0.034	.47	454	0.005	.92	454

[a]All significant correlations were considered to have a small effect.

Integration of the Feature Variables in Linear Regression Models

We found that the correlations were partially eliminated in multivariate models (Table 3). For the *active strategies* factor, only information or instructional videos had a significant effect (P=.02). The *general competency* factor showed none, and the *passive strategies* factor showed a significant effect for suggestions for planning possible activities with the family (P=.01) and diaries for documentation and development of strategies (P=.03).

XSL•FO
RenderX

Table 3. Integration of the feature variables in 3 linear regression models.

Feature variables	Active strategies		General competency		Passive strategies	
	β	*P* value	β	*P* value	β	*P* value
Individualization of app content	−.085	.21	−.042	.54	.013	.85
Fulfilling common weekly goals and tasks	.003	.97	.055	.47	.025	.74
Increasing knowledge about a healthy lifestyle	.022	.75	−.008	.91	−.026	.70
Suggestions for activities with the family	.085	.20	.005	.94	.174	.01
Diaries for documentation and development of strategies	.051	.45	.014	.84	.149	.03
Reminders for objectives	.063	.46	.152	.08	−.129	.13
Informational or instructional videos	.154	.02	.023	.73	.023	.73
Monitoring and checking progress	−.099	.11	−.086	.18	−.072	.25

Discussion

Principal Findings

The main goal of this cross-sectional study was to identify the requirements for an app that addresses stress management competencies in a cohort of highly stressed family members. We analyzed the characteristics of the target group, their individualized interests in app features and gamification aspects, their personality traits, and different types of personally defined goals.

Almost half of the study sample was identified as the high-risk population—stressed individuals with low stress competency. This underlines the importance of this study's aim. Furthermore, this group's size reflects the ever-increasing proportion of people who feel unable to effectively cope with stressors in their everyday and work–life situations, which is why the World Health Organization classified stress as the health epidemic of the 21st century and called for prevention strategies [72]. As expected, further analysis of the target group revealed a lower health index—a marker for individual health behavior based on physical activity, dietary behavior, and stress management—than that of the group with lower perceived stress levels. This finding is in line with those of prior studies, indicating that highly stressed individuals are less likely to engage in mental health promoting activities [37-40].

Notably, the target group also differed in their personally defined goals of interest for stress management. The parents who stated that they experience high stress and have low stress management competency aimed to achieve general competencies such as improvement of stress management competencies, acquisition of the ability to perform stress management techniques from anywhere, and improvement of personal resilience to stress. The nontarget group, however, aimed to achieve active strategies including performance of yoga exercises. This highlights the need to differentiate between the groups when developing and implementing mobile solutions to improve stress management competency. According to our results, the target group did not formulate specific goals but tended to have unspecific, general goals. Therefore, one might speculate that participants with subjectively higher stress levels and lower stress management competencies need more help with goal setting and more information about which strategies might reduce stress. These results further emphasize the need for tailored app features for highly stressed families. In line with Control Theory, behavior change techniques, such as goal setting, have been associated with increased intervention effects [73,74]. A study [75] evaluated a newly developed internet-based stress management intervention in a waitlist-controlled randomized trial that included principles for health behavior change such as goal setting, action planning, and coping planning for reducing stress in employees with elevated stress levels; their results showed significantly large effect differences between the intervention and waitlist control group for perceived stress at posttest. Goal-setting techniques features might thus be promising for the individual needs of stressed individuals with low stress competency.

The analysis of specific app features revealed further differences between the 2 groups regarding their app-related interests. The target participants, with low stress management competencies, indeed showed higher ratings for app features that can be used as goal-setting techniques: weekly goal and task achievements, diaries for documentation, and development of strategies and reminders for objectives. Furthermore, they had higher ratings for content individualization, connecting wearables to the app, increasing knowledge about a healthy lifestyle, suggestions for activities with the family, and informational or instructional videos. Overall, these results indicate that users identified as the high-risk population, with low stress-management competencies and high perceived stress levels, wish for features that facilitate the usage of the stress management apps. These requested features can be primarily interpreted as a need for coaching and instruction. In fact, such guided interventions have been shown to be more effective compared with unguided interventions [76]. Moreover, studies suggest guidance is conducive to the effectiveness of stress management interventions [49]. The support that might be provided in eHealth interventions can be technical or content-related, in order to ensure the correct usage [77].

A comprehensive systematic review [57] has established engagement promotion and enhancement of effectiveness as main reasons for the use of gamification in mental health promotion. In an attempt to meet this call, our study investigated the participant's interest in such elements. Nevertheless, among

XSL•FO
RenderX

gamification feature variables, only controlling and checking progress met relevance criteria. This leads to the assumption that the interest in gamification elements is mostly independent of perceived stress competencies and stress levels. The greater interest of the target group in tracking features of their progress further underpins the assumed need for coaching and instruction as well as goal setting.

Interestingly, persons reporting high stress levels and low stress management skills were younger than participants with lower self-reported stress and differed significantly in 2 personality traits variables. Specifically, their personality structure indicated lower scores in information acquisition (P<.001) and need for stimulation (P<.001). According to PSI theory, stressed individuals with low stress competency thus have more selective information acquisition and lower needs for stimulation than the nontarget group members [64]. Kuhl [78] describes individuals with selective information acquisition in accordance with Jung [79] by pointing out their analytical and structured thinking and their intuitive ability to control their behavior. Unconscious perception and behavior programs may consequently support them. A low need for stimulation, in contrast, indicates less action-oriented and more introverted behavior [64]. This is frequently associated, in other studies [80,81], with the occurrence of anxiety, stress, and depression. Thus, our results suggest that these 2 personality traits are vital for the perception and management of stress. Literature repeatedly demonstrated this influence of personality traits on stress perception [82,83]. For the development of an app to improve stress management skills, individualization of the content based on personality structure will address individuals in the target group more adequately. The content should be tailored to people with selective information acquisition and a low need for stimulation. This might be achieved by a minimalistic user interface, decreasing the participants' stimulation, thereby focusing their attention on the essential contents. Ervasti et al [84] were able to show, with a study on the influence of personality on interest in stress apps that high neuroticism levels (originated in the Big Five theory, but conceptually comparable to low stimulation need in PSI theory) were positively correlated with interest in stress management apps, which is analogous to the results of this study. The predictive value of neuroticism on stress perception was also highlighted in a comprehensive meta-analysis [58].

Furthermore, it appears that interest in some features was higher, depending on specific personally defined goals of interest. While multiple correlations were found between personally defined goals and app features, correlations were partially eliminated in multivariate models. Nevertheless, our results indicate that the probability of interest in informational or instructional videos was higher when active goals had been set. Similarly, interest in suggestions for family activities and diaries was higher when personally defined goals yielded passive strategies.

Our results support the theory that creating a stress management app requires tailored content to address the differences in perceived stress levels, stress competencies, and personality traits. These findings build on those of Lustria et al [54], who pointed out that presenting general health information without considering individual needs or personal relevance may substantially limit the extent of health behavior change. Our study substantiate this call for individualized messaging based on preassessment of key individual-difference variables by reinforcing the notion that highly vulnerable families with low perceived stress competency need an individualized app content, additionally tailoring different personality types. Tailoring works by increasing the personal relevance of health messages [85] and holds promise.

Limitations and Future Research Directions

One limitation of the study is that only individual parents were surveyed. Assessing more than one family member could not be realized at this stage of the project but will be aimed for in future studies. With respect to spill-over effects of perceived stress across family members [8-10], a holistic analysis of requirements of an app that targets the entire family is needed.

A further limitation is that the identification of our target group was only based on 2 items. This was due to practical reasons and the length of the existing survey in a larger research project. The results presented and discussed in this paper can, thus, only be regarded as exploratory and should be replicated using validated scales such as the Perceived Stress Scale [86]. Nevertheless, because of the large sample size, our results underpin those from existing research, reinforcing the notion of tailoring in the development of web-based stress management interventions.

Future research should bring these preliminary data into practice and develop and evaluate an app that adequately addresses the stress level and stress competency as well as personality traits and personal goals of the user. Furthermore, there is still no evidence to support whether already existing apps are well accepted by our target group and whether these apps provide motivating factors for long-term use to build and maintain stress management skills. For this reason, further studies on the sustainable development of apps and the support of behavior change processes, in the course of stress management, should identify the situations and conditions that can have an impact on work–life stress, coping, and goals of individual family members.

There is also a need for further research in this area to provide sustainable evidence for features and gamification elements with the aim of developing age- and gender-specific stress coping skills and, if necessary, to enable transfer to other digital media.

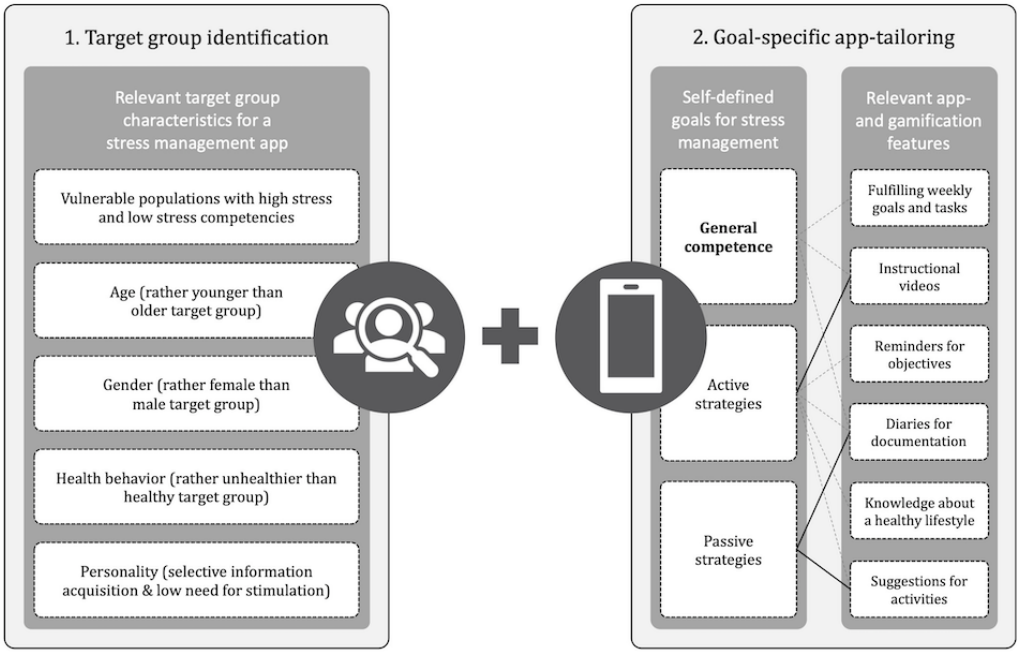
Practical Implications

Nevertheless, there are some important practical implications. Health authorities and mHealth or app developers should take our findings into account when planning and implementing tailored app-based mental health promotion interventions for families. In a first step, target group identification is necessary. The target sample—highly stressed families with low stress competencies, in this case—can be characterized by age, gender, health behavior, and personality. The design of the app, as well as its promotion, should address the unique characteristics of the target group, for example, in this case, a minimalistic user interface that decreases stimulation. In a second step, the app should be designed in such a way that individual setting of a

XSL•FO
RenderX

stress management goal is possible. These self-defined goals, defined beforehand, represent specific demands and wishes for relevant app and gamification elements (Figure 2).

Figure 2. Planned development of a tailored app to increase stress management competencies within families, based on our results. In step 2, the continuous lines depict the significant effects of the calculated linear regression models whilst the dashed lines represent significant correlations.



Conclusions

The results of this cross-sectional study show that, in order to develop an app to increase stress management competencies within families, the content should be based on preassessed of competencies, goals, and personality traits of the potential user, and thus, tailored to the user's needs. Highly stressed parents with low stress management skills want features in an app that make it easier to use and include goal setting techniques. In fact, a need for coaching and instruction was identified, which underpinned prior research showing that guided stress management interventions have more promising results.

This study delivers first results and directions to inform further research in the growing field of mobile and web-based solutions in mental health care. The relationship between integrated elements of behavior change techniques, the usage of gamification elements, and most notably, tailoring of the content of a web-based intervention and the resulting health behavior change show promise that is urgently needed with respect to increasing stress levels and its associated adverse health effects.

Acknowledgments

This study was part of a project (CASPAR Coaching App for Setting Oriented Prevention Work) supported by a German health insurance company. The views expressed in this paper are those of the authors and may not be shared by the funding agencies. The study data were analyzed independently of the study sponsors. This sponsor did not and does not play a role in the design of the study, the data analysis, the reporting of the results, or the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based questionnaire on health app tailoring requirements (translated from German).
[DOCX File , 20 KB - jmir_v23i7e26376_app1.docx]

References

1. Åkerstedt T. Psychosocial stress and impaired sleep. Scand J Work Environ Health 2006 Dec;32(6):493-501 [FREE Full text] [doi: 10.5271/sjweh.1054] [Medline: 17173205]

XSL•FO
RenderX

2. Bhatia V, Tandon RK. Stress and the gastrointestinal tract. J Gastroenterol Hepatol 2005 Mar;20(3):332-339. [doi: 10.1111/j.1440-1746.2004.03508.x] [Medline: 15740474]
3. Golden SH. A review of the evidence for a neuroendocrine link between stress, depression and diabetes mellitus. Curr Diabetes Rev 2007 Nov;3(4):252-259. [doi: 10.2174/157339907782330021] [Medline: 18220683]
4. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012 Dec 15;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
5. Hammen C. Stress and depression. Annu Rev Clin Psychol 2005;1:293-319. [doi: 10.1146/annurev.clinpsy.1.102803.143938] [Medline: 17716090]
6. Reupert A, Maybery D. Families affected by parental mental illness: a multiperspective account of issues and interventions. Am J Orthopsychiatry 2007 Jul;77(3):362-369. [doi: 10.1037/0002-9432.77.3.362] [Medline: 17696664]
7. Melchior M, Berkman LF, Niedhammer I, Zins M, Goldberg M. The mental health effects of multiple work and family demands. a prospective study of psychiatric sickness absence in the French GAZEL study. Soc Psychiatry Psychiatr Epidemiol 2007 Jul;42(7):573-582 [FREE Full text] [doi: 10.1007/s00127-007-0203-2] [Medline: 17530152]
8. Carnes AM. Bringing work stress home: the impact of role conflict and role overload on spousal marital satisfaction. J Occup Organ Psychol 2016 Dec 15;90(2):153-176 [FREE Full text] [doi: 10.1111/joop.12163]
9. Grzywacz JG, Almeida DM, McDonald DA. Work-family spillover and daily reports of work and family stress in the adult labor force. Fam Relat 2002 Aug;51(1):28-36 [FREE Full text] [doi: 10.1111/j.1741-3729.2002.00028.x]
10. Westman M. Stress and strain crossover. Hum Relat 2016 Apr 22;54(6):717-751. [doi: 10.1177/0018726701546002]
11. McMinn AM, Griffin SJ, Jones AP, van Sluijs EMF. Family and home influences on children's after-school and weekend physical activity. Eur J Public Health 2013 Oct;23(5):805-810 [FREE Full text] [doi: 10.1093/eurpub/cks160] [Medline: 23172732]
12. Arredondo EM, Elder JP, Ayala GX, Campbell N, Baquero B, Duerksen S. Is parenting style related to children's healthy eating and physical activity in Latino families? Health Educ Res 2006 Dec;21(6):862-871. [doi: 10.1093/her/cyl110] [Medline: 17032706]
13. Webb HJ, Zimmer-Gembeck MJ, Scuffham PA, Scott R, Barber B. Family stress predicts poorer dietary quality in children: examining the role of the parent-child relationship. Inf Child Dev 2018 Feb 08;27(4):e2088. [doi: 10.1002/icd.2088]
14. Smolka A, Rupp M. Die Familie als Ort der Vermittlung von Alltags- und Daseinskompetenzen. In: Harring M, Rohlfs C, Palentien C, editors. Perspektiven der Bildung: Kinder und Jugendliche in formellen, nicht-formellen und informellen Bildungsprozessen. Wiesbaden: VS Verl. für Sozialwiss; 2007:219-236 URL: https://link.springer.com/chapter/10.1007/978-3-531-90637-9_12
15. Xu H, Wen LM, Rissel C. Associations of parental influences with physical activity and screen time among young children: a systematic review. J Obes 2015;2015:546925 [FREE Full text] [doi: 10.1155/2015/546925] [Medline: 25874123]
16. Borrmann A, Mensink GBM, KiGGS Study Group. [Fruit and vegetable consumption by children and adolescents in Germany: results of KiGGS wave 1]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Sep;58(9):1005-1014. [doi: 10.1007/s00103-015-2208-4] [Medline: 26141246]
17. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014 Aug 1;129 Suppl 2(12):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
18. Geene R, Boger M. Literatur- und Datenbankrecherche zu Gesundheitsförderungs- und Präventionsansätzen bei Alleinerziehenden und Auswertung der vorliegenden Evidenz. GKV-Spitzenverband. 2017. URL: https://www.gkv-buendnis.de/fileadmin/user_upload/Literaturrecherche_Alleinerziehende_Geene_2017.pdf [accessed 2020-11-30]
19. Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clin Psychol (New York) 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
20. Fjorback LO, Arendt M, Ornbøl E, Fink P, Walach H. Mindfulness-based stress reduction and mindfulness-based cognitive therapy: a systematic review of randomized controlled trials. Acta Psychiatr Scand 2011 Aug;124(2):102-119. [doi: 10.1111/j.1600-0447.2011.01704.x] [Medline: 21534932]
21. Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits. a meta-analysis. J Psychosom Res 2004 Jul;57(1):35-43. [doi: 10.1016/S0022-3999(03)00573-7] [Medline: 15256293]
22. Querstret D, Morison L, Dickinson S, Cropley M, John M. Mindfulness-based stress reduction and mindfulness-based cognitive therapy for psychological health and well-being in nonclinical samples: a systematic review and meta-analysis. Int J Stress Manag 2020 Nov;27(4):394-411. [doi: 10.1037/str0000165]
23. Avila-Palencia I, de Nazelle A, Cole-Hunter T, Donaire-Gonzalez D, Jerrett M, Rodriguez DA, et al. The relationship between bicycle commuting and perceived stress: a cross-sectional study. BMJ Open 2017 Jun 23;7(6):e013542 [FREE Full text] [doi: 10.1136/bmjopen-2016-013542] [Medline: 28645948]
24. Bischoff LL, Otto A, Hold C, Wollesen B. The effect of physical activity interventions on occupational stress for health personnel: a systematic review. Int J Nurs Stud 2019 Sep;97:94-104. [doi: 10.1016/j.ijnurstu.2019.06.002] [Medline: 31234106]

XSL•FO
RenderX

25. von Haaren B, Ottenbacher J, Muenz J, Neumann R, Boes K, Ebner-Priemer U. Does a 20-week aerobic exercise training programme increase our capabilities to buffer real-life stressors? a randomized, controlled trial using ambulatory assessment. Eur J Appl Physiol 2016 Feb;116(2):383-394. [doi: 10.1007/s00421-015-3284-8] [Medline: 26582310]
26. Sloan RP, Shapiro PA, DeMeersman RE, Bagiella E, Brondolo EN, McKinley PS, et al. Impact of aerobic training on cardiovascular reactivity to and recovery from challenge. Psychosom Med 2011;73(2):134-141 [FREE Full text] [doi: 10.1097/PSY.0b013e31820a1174] [Medline: 21257979]
27. Klaperski S, von Dawans B, Heinrichs M, Fuchs R. Effects of a 12-week endurance training program on the physiological response to psychosocial stress in men: a randomized controlled trial. J Behav Med 2014 Dec;37(6):1118-1133. [doi: 10.1007/s10865-014-9562-9] [Medline: 24659155]
28. Ma X, Yue Z, Gong Z, Zhang H, Duan N, Shi Y, et al. The effect of diaphragmatic breathing on attention, negative affect and stress in healthy adults. Front Psychol 2017;8:874 [FREE Full text] [doi: 10.3389/fpsyg.2017.00874] [Medline: 28626434]
29. Hopper SI, Murray SL, Ferrara LR, Singleton JK. Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review. JBI Database System Rev Implement Rep 2019 Sep;17(9):1855-1876. [doi: 10.11124/JBISRIR-2017-003848] [Medline: 31436595]
30. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: a meta-analysis. J Psychosom Res 2015 Jun;78(6):519-528. [doi: 10.1016/j.jpsychores.2015.03.009] [Medline: 25818837]
31. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
32. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
33. Oken BS, Wahbeh H, Goodrich E, Klee D, Memmott T, Miller M, et al. Meditation in stressed older adults: improvements in self-rated mental health not paralleled by improvements in cognitive function or physiological measures. Mindfulness (N Y) 2017 Jun;8(3):627-638 [FREE Full text] [doi: 10.1007/s12671-016-0640-7] [Medline: 28603562]
34. Hussain JN, Greaves RF, Cohen MM. A hot topic for health: results of the global sauna survey. Complement Ther Med 2019 Jun;44:223-234. [doi: 10.1016/j.ctim.2019.03.012] [Medline: 31126560]
35. Antonelli M, Barbieri G, Donelli D. Effects of forest bathing (shinrin-yoku) on levels of cortisol as a stress biomarker: a systematic review and meta-analysis. Int J Biometeorol 2019 Aug;63(8):1117-1134. [doi: 10.1007/s00484-019-01717-x] [Medline: 31001682]
36. Hunter MR, Gillespie BW, Chen SY. Urban nature experiences reduce stress in the context of daily life based on salivary biomarkers. Front Psychol 2019;10:722 [FREE Full text] [doi: 10.3389/fpsyg.2019.00722] [Medline: 31019479]
37. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med 2014 Jan;44(1):81-121 [FREE Full text] [doi: 10.1007/s40279-013-0090-5] [Medline: 24030837]
38. Schultchen D, Reichenberger J, Mittl T, Weh TRM, Smyth JM, Blechert J, et al. Bidirectional relationship of stress and affect with physical activity and healthy eating. Br J Health Psychol 2019 May;24(2):315-333 [FREE Full text] [doi: 10.1111/bjhp.12355] [Medline: 30672069]
39. Ng DM, Jeffery RW. Relationships between perceived stress and health behaviors in a sample of working adults. Health Psychol 2003 Nov;22(6):638-642. [doi: 10.1037/0278-6133.22.6.638] [Medline: 14640862]
40. Rao KU, Russel RW. Effects of stress on goal setting behavior. J Abnorm Soc Psychol 1960 Nov;61:380-388. [doi: 10.1037/h0040176] [Medline: 13739530]
41. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
42. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
43. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
44. Lanata A, Valenza G, Nardelli M, Gentili C, Scilingo EP. Complexity index from a personalized wearable monitoring system for assessing remission in mental health. IEEE J Biomed Health Inform 2015 Jan;19(1):132-139. [doi: 10.1109/JBHI.2014.2360711] [Medline: 25291802]
45. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343 [FREE Full text] [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
46. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]

XSL•FO
RenderX

47. Cuijpers P, Kleiboer A, Karyotaki E, Riper H. Internet and mobile interventions for depression: opportunities and challenges. Depress Anxiety 2017 Jul;34(7):596-602. [doi: 10.1002/da.22641] [Medline: 28471479]
48. Cheng SK, Dizon J. Computerised cognitive behavioural therapy for insomnia: a systematic review and meta-analysis. Psychother Psychosom 2012;81(4):206-216. [doi: 10.1159/000335379] [Medline: 22585048]
49. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
50. Beintner I, Jacobi C, Taylor CB. Participant adherence to the internet-based prevention program StudentBodies™ for eating disorders — a review. Internet Interventions 2014 Mar;1(1):26-32. [doi: 10.1016/j.invent.2014.03.001]
51. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
52. Lustria MLA, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Mar;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
53. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Mar;15(1):1-4. [doi: 10.1093/her/15.1.1] [Medline: 10788196]
54. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
55. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
56. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
57. Cheng V, Davenport T, Johnson D, Vella K, Hickie I. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
58. Connor-Smith JK, Flachsbart C. Relations between personality and coping: a meta-analysis. J Pers Soc Psychol 2007 Dec;93(6):1080-1107. [doi: 10.1037/0022-3514.93.6.1080] [Medline: 18072856]
59. Vollrath M, Torgersen S. Personality types and coping. Pers Individ Dif 2000 Aug;29(2):367-378. [doi: 10.1016/s0191-8869(99)00199-3]
60. Stachl C, Hilbert S, Au J, Buschek D, De Luca A, Bischl B, et al. Personality traits predict smartphone usage. Eur J Pers 2017 Aug 02;31(6):701-722 [FREE Full text] [doi: 10.1002/per.2113]
61. Su J, Dugas M, Guo X, Gao GG. Influence of personality on mhealth use in patients with diabetes: prospective pilot study. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e17709 [FREE Full text] [doi: 10.2196/17709] [Medline: 32773382]
62. Ghaban W, Hendley R. How different personalities benefit from gamification. Interact Comput 2019 Apr 08;31(2):138-153. [doi: 10.1093/iwc/iwz009]
63. Kuhl J. The volitional basis of personality systems interaction theory. Int J Educ Res 2000 Jan;33(7-8):665-703 [FREE Full text] [doi: 10.1016/s0883-0355(00)00045-8]
64. Scheffer D, Loerwald D, Mainz D. Messung von impliziten Persönlichkeits-Systemen mit Hilfe der visuellen Testmethode des Visual Questionnaire ViQ. In: Arbeitspapiere der Nordakademie. Elmshorn: Nordakademie - Hochschule der Wirtschaft; Mar 2009.
65. Meixner C, Baumann H, Wollesen B. Personality traits, gamification and features to develop an app to reduce physical inactivity. Information 2020 Jul 19;11(7):367. [doi: 10.3390/info11070367]
66. Questback GmbH. EFS survey. Questback. Cologne URL: https://www.questback.com/de/online-befragungstool/ [accessed 2020-11-30]
67. Hagger M, Keatley D, Chan D. CALO-RE taxonomy of behavior change techniques. In: Inklund RC, Tenenbaum G, editors. Encyclopedia of Sport and Exercise Psychology. Thousand Oaks: SAGE Publications, Inc; 2014:100-104.
68. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982 Nov;36(5):936-942. [doi: 10.1093/ajcn/36.5.936] [Medline: 7137077]
69. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; 2010:2014-2092.
70. 10 guidelines of the German Nutrition Society (DGE) for a wholesome diet. Deutsche Gesellschaft für Ernährung e.V. URL: https://www.dge.de/ernaehrungspraxis/vollwertige-ernaehrung/10-regeln-der-dge/en/ [accessed 2020-12-07]
71. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
72. The Global Burden of Disease 2004 Update. Geneva: World Health Organization; 2008.
73. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

XSL•FO
RenderX

74. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995 May;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]
75. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
76. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
77. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
78. Kuhl J. Motivation und Persönlichkeit: Interaktionen psychischer Systeme. Göttingen: Hogrefe Verl. für Psychologie; 2001.
79. Jung C. Psychologische Typen. Solothurn: Walter-Verl; 2006.
80. Uliaszek AA, Zinbarg RE, Mineka S, Craske MG, Sutton JM, Griffith JW, et al. The role of neuroticism and extraversion in the stress-anxiety and stress-depression relationships. Anxiety Stress Coping 2010 Jul;23(4):363-381 [FREE Full text] [doi: 10.1080/10615800903377264] [Medline: 19890753]
81. Jylhä P, Melartin T, Isometsä E. Relationships of neuroticism and extraversion with axis I and II comorbidity among patients with DSM-IV major depressive disorder. J Affect Disord 2009 Apr;114(1-3):110-121. [doi: 10.1016/j.jad.2008.06.011] [Medline: 18687471]
82. Cramer C, Binder K. Zusammenhänge von Persönlichkeitsmerkmalen und Beanspruchungserleben im Lehramt. Ein internationales systematisches Review. Z Erziehungswiss 2015 Jan 31;18(1):101-123. [doi: 10.1007/s11618-014-0605-3]
83. Kim LE, Jörg V, Klassen RM. A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. Educ Psychol Rev 2019;31(1):163-195 [FREE Full text] [doi: 10.1007/s10648-018-9458-2] [Medline: 30930595]
84. Ervasti M, Kallio J, Määttänen I, Mäntyjärvi J, Jokela M. Influence of personality and differences in stress processing among Finnish students on interest to use a mobile stress management app: survey study. JMIR Ment Health 2019 May 13;6(5):e10039 [FREE Full text] [doi: 10.2196/10039] [Medline: 31094358]
85. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
86. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

Abbreviations

mHealth: mobile health
PSI: Personality System Interaction

Edited by R Kukafka; submitted 09.12.20; peer-reviewed by J Kallio, N Fijacko; comments to author 17.02.21; revised version received 24.03.21; accepted 04.05.21; published 30.07.21.

Please cite as:
Bischoff LL, Baumann H, Meixner C, Nixon P, Wollesen B
App-Tailoring Requirements to Increase Stress Management Competencies Within Families: Cross-sectional Survey Study
J Med Internet Res 2021;23(7):e26376
URL: https://www.jmir.org/2021/7/e26376
doi:10.2196/26376
PMID:34328439

©Laura Luise Bischoff, Hannes Baumann, Charlotte Meixner, Patricia Nixon, Bettina Wollesen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Self-guided Cognitive Behavioral Therapy Apps for Depression: Systematic Assessment of Features, Functionality, and Congruence With Evidence

Laura Martinengo[1], MD; Anne-Claire Stona[1], MD; Konstadina Griva[1], PhD; Paola Dazzan[2,3], PhD; Carmine Maria Pariante[4], MD, PhD; Florian von Wangenheim[5], PhD; Josip Car[1,6], MD, PhD

[1]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[3]National Institute for Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, London, United Kingdom
[4]Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, London, United Kingdom
[5]Professor of Technology Marketing, Department of Management, Technology & Economics, ETH Zurich, Zurich, Switzerland
[6]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Josip Car, MD, PhD
Centre for Population Health Sciences
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
Clinical Sciences Building, Level 18
11 Mandalay Road
Singapore
Singapore
Phone: 65 6904 7005
Email: josip.car@ntu.edu.sg

Abstract

Background: Mental health disorders affect 1 in 10 people globally, of whom approximately 300 million are affected by depression. At least half of the people affected by depression remain untreated. Although cognitive behavioral therapy (CBT) is an effective treatment, access to mental health specialists, habitually challenging, has worsened because of the COVID-19 pandemic. Internet-based CBT is an effective and feasible strategy to increase access to treatment for people with depression. Mental health apps may further assist in facilitating self-management for people affected by depression; however, accessing the correct app may be cumbersome given the large number and wide variety of apps offered by public app marketplaces.

Objective: This study aims to systematically assess the features, functionality, data security, and congruence with evidence of self-guided CBT-based apps targeting users affected by depression that are available in major app stores.

Methods: We conducted a systematic assessment of self-guided CBT-based apps available in Google Play and the Apple App Store. Apps launched or updated since August 2018 were identified through a systematic search in the 42matters database using CBT-related terms. Apps meeting the inclusion criteria were downloaded and assessed using a Samsung Galaxy J7 Pro (Android 9) and iPhone 7 (iOS 13.3.1). Apps were appraised using a 182-question checklist developed by the research team, assessing their general characteristics, technical aspects and quality assurance, and CBT-related features, including 6 evidence-based CBT techniques (ie, psychoeducation, behavioral activation, cognitive restructuring, problem solving, relaxation, and exposure for comorbid anxiety) as informed by a CBT manual, CBT competence framework, and a literature review of internet-based CBT clinical trial protocols. The results were reported as a narrative review using descriptive statistics.

Results: The initial search yielded 3006 apps, of which 98 met the inclusion criteria and were systematically assessed. There were 20 well-being apps; 65 mental health apps, targeting two or more common mental health disorders, including depression; and 13 depression apps. A total of 28 apps offered at least four evidence-based CBT techniques, particularly depression apps. Cognitive restructuring was the most common technique, offered by 79% (77/98) of the apps. Only one-third of the apps offered suicide risk management resources, whereas 17% (17/98) of the apps offered COVID-19–related information. Although most apps included a privacy policy, only a third of the apps presented it before account creation. In total, 82% (74/90) of privacy

XSL•FO
RenderX

policies stated sharing data with third-party service providers. Half of the app development teams included academic institutions or health care providers.

Conclusions: Only a few self-guided CBT-based apps offer comprehensive CBT programs or suicide risk management resources. Sharing of users' data is widespread, highlighting shortcomings in health app market governance. To fulfill their potential, self-guided CBT-based apps should follow evidence-based clinical guidelines, be patient centered, and enhance users' data security.

(J Med Internet Res 2021;23(7):e27619) doi:10.2196/27619

KEYWORDS

cognitive behavioral therapy; CBT; depression; mobile applications; apps; telemedicine; mHealth; self-guided CBT-based apps; self-management; mobile phone

Introduction

Background

Mental health disorders affect 1 in 10 people globally [1], of which approximately 300 million are affected by depression [2]. At least half of the affected people remain untreated [3-5] because of many reasons [6-8], including low perceived need for treatment and self-reliance [7,8], stigma [9], difficulty accessing specialist care [10], fragmentation of the mental health care system [11], and high costs of treatment [12]. Access to psychiatrists and clinical psychologists is challenging, particularly in low- and middle-income countries [13] and for people of lower socioeconomic status, poorer health, and rural and hard-to-reach communities in high-income economies [14,15]. Since the onset of the COVID-19 pandemic, these inadequacies have been exacerbated by physical isolation, and fear and uncertainty [16,17] led to a sudden increase in mental health concerns [18].

Cognitive behavioral therapy (CBT), one of the most widely used and researched forms of psychotherapy [19], is effective in treating a wide range of mental disorders, including depression [20-22]. CBT is a "structured, time-limited, present-focused" [23] therapy that promotes behavioral adaptation by coaching users to use a variety of tailored cognitive and behavioral techniques [24]. Traditional face-to-face CBT is costly, time-consuming, and subject to the availability of trained providers [10], limiting accessibility. To increase access to therapy, a number of internet-based CBT (iCBT) programs [25-27] offer an acceptable and effective alternative [28-30]. Therapist-guided iCBT has been shown to be as effective as face-to-face sessions [28], whereas self-guided interventions, albeit effective, are not at par with guided web-based programs [30-32].

Over the last decade, there has been increasing interest in using smartphone apps to support mental health disorder management and well-being [33], which has been further intensified since the COVID-19 pandemic [34,35]. Recent reports indicate that nearly 320,000 health apps [36] are available in major app stores, of which more than 10,000 are mental health apps [37]. Smartphone penetration is increasing worldwide, particularly in high-income countries [38], making apps a useful means to improve access to care for people unable or unwilling to consult a health care provider by offering interventions to be used at the time and place of the user's convenience at low cost [39,40].

Apps can improve patient activation and disease self-management by increasing access to information and facilitating communication with health care providers and peers [41]. Two recent reviews, comprising one systematic review and one umbrella review [42,43], concluded that apps were effective in improving users' emotional symptoms, including depression, although effect sizes varied according to the intervention and comparator groups, as well as the study quality. However, several shortcomings of mental health apps have been repeatedly described, including substantial dropout rates [44], mishandling of users' personal health information [45], poor app credibility, and lack of content personalization [41]. In addition, most health apps available in major app stores are not evidence based and have not been validated in clinical trials or approved by regulatory agencies [37,46], underlining concerns about the effectiveness and safety of publicly available apps.

Previous work on self-guided CBT apps for depression noted only a few apps offering some evidence-based techniques, with limited user engagement features and a dominant focus solely on depression [47,48]. However, people with depression searching for an app will often retrieve a much wider variety of apps, including apps targeting several mental health disorders or those aimed at improving general well-being for healthy individuals. Therefore, we considered it important to conduct an updated assessment of self-guided CBT-based apps available in Google Play and Apple App Store encompassing well-being, general mental health, and depression apps to more closely align with real-life users' experience.

Objective

This study aims to systematically assess features, functionality, and congruence with evidence of self-guided CBT-based apps available in major app stores, targeting users affected by depression.

Methods

Overview

We developed a rigorous systematic assessment based on systematic literature review methodology, as used in several other studies [49-52]. The process included a systematic search for apps, two-step selection, and assessment following an exhaustive set of criteria developed by the research team.

XSL•FO
RenderX

Development of Assessment Criteria

The assessment criteria aimed to provide an in-depth and holistic analysis of the apps, including clinical and technical characteristics. CBT-related features constitute a fundamental aspect of the overall appraisal, particularly the assessed apps' content compliance with evidence-based practice, as all health apps should ideally be based on sound evidence. Technical characteristics were included in the assessment if they were available in the app, app store description, or associated website and were not dependent on the assessor's subjective opinion or required access to the app's back-end data. The form was divided into three sections:

1. *General characteristics* extracted from the app store description, including developer, platform, app version number, category and ratings, number of downloads, cost, country of origin, languages featured in the app, target group, scope of the app, and therapy modalities offered by the app.
2. *CBT-related features*, based on *Cognitive Behavior Therapy: Basics and Beyond* by Judith Beck [53], a CBT competence framework [54] developed by the Improving Access to Psychological Therapies program in the United Kingdom, and a literature review of iCBT clinical trial protocols to identify distinct characteristics relevant to digital interventions. They comprised the following (Multimedia Appendix 1):
   - *Evidence-based CBT techniques*, as described in the reference sources mentioned earlier, are routinely used in face-to-face practice: *psychoeducation*, about depression and CBT;*behavioral activation (BA)*, including activity and task scheduling, suggestion of pleasurable activities, and monitoring of completed activities; *cognitive restructuring*, including assessment of automatic thoughts and core beliefs using thought records and completing behavioral experiments to challenge automatic thoughts and core beliefs; and *other techniques* frequently used in CBT, including *problem solving*, *relaxation*, and *exposure techniques* for comorbid anxiety.
   - *Procedures related to the structure of face-to-face CBT sessions*: CBT sessions are highly structured and offer users strategies to cope after the end of therapy. We assessed whether relevant components of CBT sessions were offered, including structured, modular sessions, mood monitoring, suicide risk assessment, goal setting, homework assignment, *therapeutic alliance* [55], and coping strategies after completing the program. Following Tremain et al [55], we conceptualized the *therapeutic alliance* for self-guided interventions as strategies that encourage users' engagement with the app and adherence to assigned tasks, such as gamification, notifications, and reminders.
   - *Other information*, including access to professional advice for distressed users, information related to the mental health impact of the COVID-19 pandemic as a proxy for continuous improvement of app content, and other techniques not mentioned elsewhere.
3. *Technical aspects and quality assurance of the app*, based on an assessment framework developed by our center [51,52], comprising ease-of-use, app credibility, presence of advertisements, privacy and security safeguards, including thoroughness of the privacy policy, privacy settings, and authentication to access the app content, among others. App credibility included appropriately referenced app content, qualifications of the app development team, presence of *information does not replace provider's advice* disclaimers, and published evidence of app effectiveness, assessed by searching PubMed for publications using the app name as keyword.

App Selection

App Search

A systematic search for apps available in the Apple App Store and Google Play Store was performed on February 19, 2020, in 42matters [56], a proprietary database, using the search terms *cognitive behavioral therapy*, *cognitive behavioural therapy*, *cognitive therapy*, *CBT*, *behavioral therapy*, *behavioural therapy*, *behavioral activation*, *behavioural activation*, *online therapy*, *psychotherapy*, *counselling*, and *talking therapy*. A total of 4 app store categories were included: education, health and fitness, lifestyle, and medical. For apps available in only one app store, a web search was performed to look for the other version, and if available, it was downloaded and assessed.

Eligibility Criteria

We included apps described as based on CBT, exclusively or associated to other psychotherapeutic modalities; offering CBT-based activities, exclusively or associated to other psychotherapeutic modalities; aiming to improve mood and well-being for people with low mood or depression and targeting depression alone or associated with another mental disorder; uploaded or updated from August 1, 2018, as regular updates of an app seem to directly affect app quality [57]; free, freemium (app is free to download but requires payment to activate extra features), or paid; available for download in the Apple App Store or Google Play Store; and in English.

We excluded apps offering CBT for other mental health disorders (ie, standalone anxiety, insomnia, posttraumatic stress disorder, etc); offering non–CBT-based mood recording or journaling, as referred to in the app store description; offering health care provider– or counselor-guided CBT modules, targeting health care providers or caregivers of a person affected by depression; offering teleconsultation services with physicians, psychologists, counselors, or other health care providers; removed from the app stores at the time of download, requiring a sign-up code provided by an institution, or inaccessible after two attempts due to technical problems; and in a language other than English.

App Assessment

App selection followed a systematic, two-step process. We first screened the app title and app store description from the 42matters search output and downloaded all apps included in the first step to assess eligibility. All assessments were performed using a Samsung Galaxy J7 Pro (Android 9) and

XSL•FO
RenderX

iPhone 7 (iOS 13.3.1). If apps were available in both app stores, both versions were assessed to account for any difference in functionalities, and each version was included and counted separately. Individual apps belonging to a suite of related apps were combined and counted as one app.

Eligible apps were then assessed by the researcher (a medical doctor) for all available CBT techniques. If the app presented a modular intervention, it was used repeatedly to complete all modules. For consistency of assessments, a user persona was developed outlining demographic information, personal and medical history, responses to assessment questionnaires, and opening statements for conversational agent dialogs.

Data Analysis

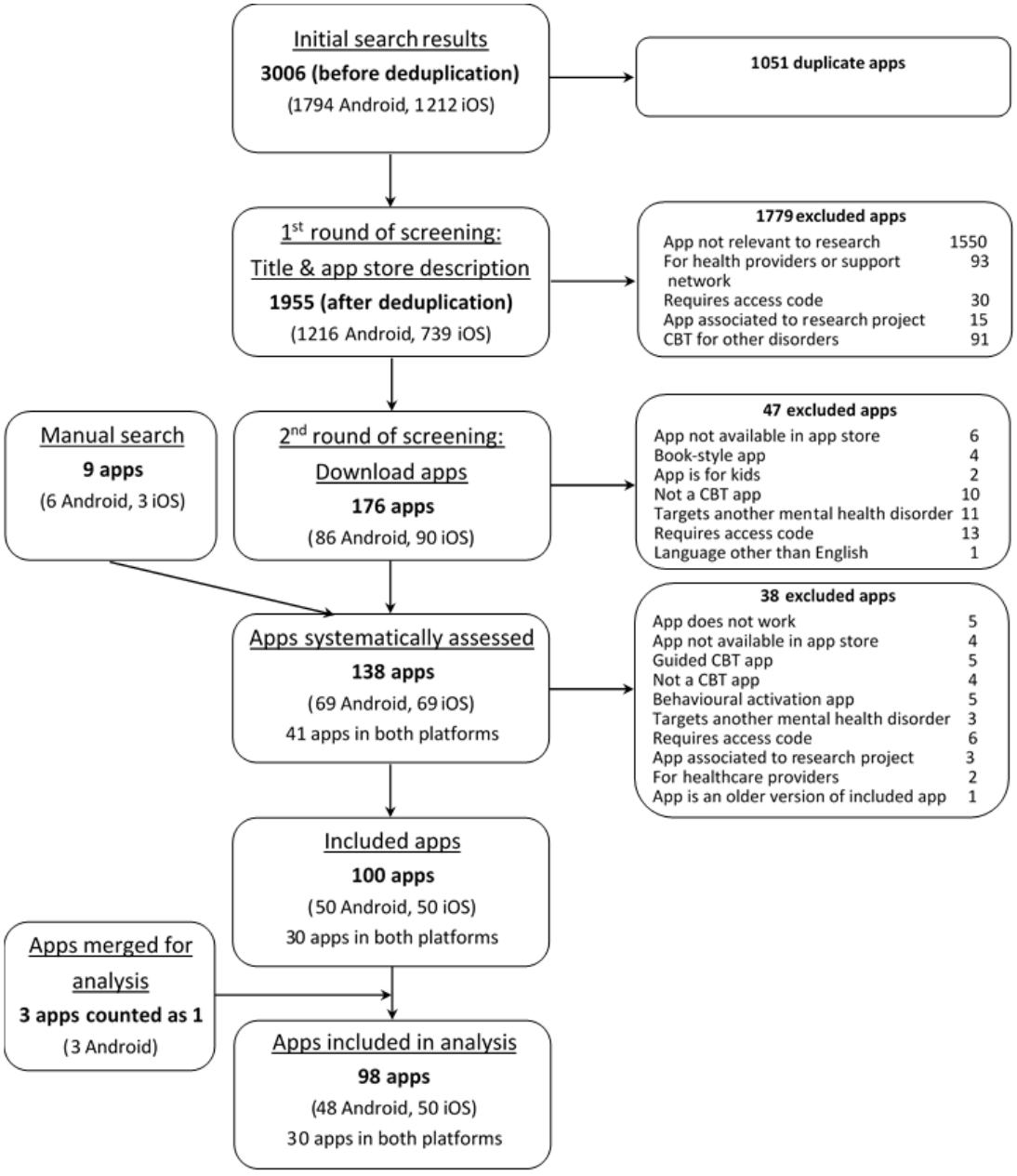
Descriptive statistics were used to analyze the data. The study results were tabulated and reported as a narrative synthesis. Data extraction and analysis were performed using Microsoft Excel.

Results

Overview

The keyword search retrieved 1955 results after duplicates were removed, of which 140 apps were downloaded and 100 apps were included in this analysis. A total of 3 apps (ie, eGuru Depression [58], eGuru Mood Diary [59], and eGuru Thought Diary [60]) belonged to a suite of related apps and were analyzed together as one app. Therefore, 98 apps were finally included in the analysis. The app search and selection processes are summarized in Figure 1. Multimedia Appendix 2 presents the list of assessed apps and a summary of CBT-related features.

Figure 1. App selection flowchart. CBT: cognitive behavioral therapy.



XSL•FO
RenderX

General Characteristics of Apps

Table 1 summarizes the characteristics of the included apps. Apps were grouped into three distinct categories: (1) *well-being apps*, providing CBT-based activities to improve users' general well-being; (2) *mental health apps*, offering self-guided activities to manage two or more common mental health disorders, including depression; and (3) *depression apps* offering self-guided CBT exclusively for depression.

There were 48 Android apps and 50 iOS apps. A total of 30 apps were available on both platforms, 18 were Android-only apps and 20 were iOS-only apps. A total of 74% (73/98) of apps belonged to the health and fitness app store category, whereas 19% (19/98) of apps were listed as medical apps. A total of 27% (13/48) of Android apps were downloaded more than 100,000 times, including 11 mental health apps, 1 depression app, and 1 well-being app. A total of 4 mental health apps (ie, Moodpath [61], Sanvello [62], Wysa [63], and Youper [64]) and 1 well-being app (ie, Reflectly [65]) were downloaded more than 1,000,000 times. Three-quarter (n=73) of the apps had users' ratings above 3.5 on a 1-5 scale and 42% (41/98) of apps had ratings above 4.5. Apps were developed in 16 different countries (Table 2).

XSL•FO
RenderX

Table 1. General characteristics of the included apps (N=98).

Feature	App category, n (%)			Total (N=98), n (%)
	Well-being (n=20)	Mental health (n=65)	Depression (n=13)	
App store category				
Education	1 (5)	0 (0)	0 (0)	1 (1)
Health and fitness	19 (95)	46 (71)	8 (62)	73 (74)
Lifestyle	0 (0)	5 (8)	0 (0)	5 (5)
Medical	0 (0)	14 (22)	5 (38)	19 (19)
App store rating				
3.6 star to 5 star	15 (75)	52 (80)	6 (46)	73 (74)
1 star to 3.5 star	0 (0)	7 (11)	2 (15)	9 (9)
No ratings	5 (25)	6 (9)	5 (38)	16 (16)
App cost				
Free	8 (40)	27 (42)	4 (31)	39 (40)
Free with in-app purchase	11 (55)	28 (43)	9 (69)	48 (49)
Paid	1 (5)	10 (15)	0 (0)	11 (11)
Language				
English	16 (80)	47 (72)	12 (92)	75 (77)
English and other languages	4 (20)	18 (28)	1 (8)	23 (23)
Target user of the app				
No target user	20 (100)	65 (100)	11 (85)	96 (98)
Youth (ages 12-18 years)	0 (0)	0 (0)	2 (15)	2 (2)
Psychotherapy modalities used by the app				
CBT[a]	8 (40)	31 (48)	9 (69)	48 (49)
CBT and other modalities	12 (60)	34 (52)	4 (31)	50 (51)
Number of evidence-based CBT techniques offered by the app				
0	5 (25)	4 (6)	0 (0)	9 (9)
1	3 (15)	6 (9)	1 (8)	10 (10)
2	11 (55)	25 (34)	2 (15)	38 (39)
3	1 (10)	10 (20)	2 (15)	13 (13)
4	0 (0)	16 (23)	8 (62)	24 (24)
5	0 (0)	4 (8)	0 (0)	4 (4)
6	0 (0)	0 (0)	0 (0)	0 (0)
Emergency contact information for users at risk of suicide				
Yes	2 (10)	23 (35)	8 (62)	33 (34)
No	18 (90)	42 (65)	5 (38)	65 (66)
Information related to COVID-19 pandemic				
Yes	0 (0)	16 (25)	1 (8)	17 (17)
No	20 (100)	49 (75)	12 (92)	81 (83)
App works as intended[b]				
Yes	18 (90)	62 (95)	12 (92)	92 (94)
No	2 (10)	3 (5)	1 (8)	6 (6)

[a]CBT: cognitive behavioral therapy.

XSL•FO
RenderX

[b]Did not malfunction or crash during usage.

Table 2. List of countries where the self-guided cognitive behavioral therapy–based apps were developed (N=98).

Country	App category, n (%)			Total, n (%)
	Well-being (n=20)	Mental health (n=65)	Depression (n=13)	
Australia	0 (0)	4 (6)	2 (15)	6 (6)
Bulgaria	2 (10)	0 (0)	0 (0)	2 (2)
Canada	1 (5)	1 (2)	3 (23)	5 (5)
Colombia	0 (0)	1 (2)	0 (0)	1 (1)
Cyprus and Belarus	1 (5)	0 (0)	0 (0)	1 (1)
Denmark	2 (10)	0 (0)	0 (0)	2 (2)
Germany	0 (0)	4 (6)	0 (0)	4 (4)
India	2 (10)	4 (6)	1 (8)	7 (7)
Israel	0 (0)	2 (3)	0 (0)	2 (2)
New Zealand	0 (0)	2 (3)	0 (0)	2 (2)
Norway	0 (0)	2 (3)	1 (8)	3 (3)
Poland	2 (10)	3 (5)	0 (0)	5 (5)
Sweden	2 (10)	0 (0)	0 (0)	2 (2)
The Netherlands	0 (0)	1 (2)	0 (0)	1 (1)
United Kingdom	0 (0)	10 (15)	2 (15)	12 (12)
United States	6 (30)	28 (43)	4 (31)	38 (39)
Unknown	2 (10)	3 (5)	0 (0)	5 (5)

A total of 25% (16/65) of mental health apps and 8% (1/13) of depression apps offered advice to manage the uncertainties and anxiety associated with COVID-19. In addition, 2 apps offered their subscription-only content for free to support users during the pandemic.

CBT-Related Features

Overview

Apps included a wide variety of CBT-related features, from structured CBT modules resembling face-to-face interventions to journaling applications, presenting guided questionnaires loosely based on the cognitive model. Table 3 outlines the CBT-related features offered by the apps.

XSL•FO
RenderX

Table 3. Cognitive behavioral therapy–related techniques offered by the apps (N=98).

Feature	App category, n (%)			Total (N=98), n (%)
	Well-being (n=20)	Mental health (n=65)	Depression (n=13)	
Evidence-based CBT[a] techniques offered by the apps				
Psychoeducation	2 (10)	44 (68)	11 (85)	57 (58)
Behavioral activation	2 (10)	28 (43)	11 (85)	41 (42)
Cognitive restructuring	10 (50)	54 (83)	13 (100)	77 (79)
Problem solving	4 (20)	6 (9)	2 (15)	12 (12)
Relaxation	10 (50)	38 (38)	6 (46)	54 (55)
Exposure	1 (5)	4 (6)	0 (0)	5 (5)
Procedures related to the structure of CBT sessions				
Content offered in modules	2 (10)	12 (18)	8 (62)	22 (22)
Current mood monitoring	14 (70)	45 (69)	7 (54)	66 (67)
App administers screening questionnaire	0 (0)	21 (32)	11 (85)	32 (33)
Suicide risk management	2 (10)	23 (35)	9 (69)	34 (35)
Homework assignment	0 (0)	16 (25)	4 (31)	20 (20)
Therapeutic alliance	14 (70)	29 (45)	3 (23)	46 (47)
Strategies to cope after completing the modules	2 (10)	3 (5)	4 (31)	9 (9)
Other functionalities				
Journaling	3 (15)	7 (11)	0 (0)	10 (10)
Gratitude or positive thought records	2 (10)	10 (15)	0 (0)	12 (12)
Forums	2 (10)	4 (6)	0 (0)	6 (6)
Web-based games	1 (5)	4 (6)	0 (0)	6 (6)
Others	1 (5)	18 (28)	5 (38)	24 (25)

[a]CBT: cognitive behavioral therapy.

Evidence-Based CBT Techniques

Overview

Most well-being apps (19/20, 95%) and over half of the mental health apps (35/65, 54%) offered up to two evidence-based CBT techniques, whereas most depression apps (10/13, 77%) offered three or four evidence-based techniques. Table 3 presents a summary of the CBT techniques assessed. A total of 25% (5/20) of well-being apps and 6% (4/65) of mental health apps offered questionnaires or journaling templates based on the cognitive model but no evidence-based techniques.

Psychoeducation

Psychoeducation offered by the apps was mostly related to aspects of CBT, that is, explanations about cognitive distortions assessed in the app, automatic thoughts, and the CBT model. Information about depression was found in over a third of mental health apps (15/44, 34%) and most of the depression apps (10/11, 91%).

Behavioral Activation

A total of 42% (41/98) of apps offered BA techniques or modules frequently assisting users to track (37/41, 90%) or schedule (30/41, 73%) their activities, whereas 56% (23/41) of apps suggested activities to engage in.

Cognitive Restructuring

Overall, 79% (77/98) of apps of the apps offered cognitive restructuring. Most apps (73/77, 95%) guided users to identify automatic thoughts, whereas approximately two-third of the apps offered guidance to reframe automatic thoughts (55/77, 71%) and identify the emotion linked to automatic thoughts (50/77, 65%) and/or cognitive distortions associated with automatic thoughts (48/77, 62%). Approximately two-third of the apps (50/77, 65%) obtained information using a template of follow-up questions similar to a thought record chart. A total of 6 apps (6/77, 8%) offered a noninteractive thought record template as their only feature.

Problem Solving

Problem-solving techniques or modules were offered by 12% (12/98) of apps, consisting of templates to plan a series of steps to solve a given problem.

Relaxation

A total of 54 apps offered relaxation modules. They consisted of mindfulness or meditation (33/54, 61%) apps or breathing exercise (31/54, 57%) apps. In total, 85% (46/54) of apps developed their own relaxation modules, whereas the rest offered links to other apps or websites.

XSL•FO
RenderX

Exposure Techniques for Comorbid Anxiety

In total, 5% (1/20) of well-being apps and 6% (4/65) of mental health apps included techniques to expose users to anxiety-provoking stimuli. A total of 3 apps requested the user to record the exposure exercises, whereas 2 apps offered a reward for completed tasks.

Procedures Related to the Structure of Face-to-Face CBT Sessions

Overview

A total of 22 apps offered content in structured modules. These modules resembled face-to-face CBT sessions in the 3 depression apps. Furthermore, 9 apps offered advice on how to cope at the end of the modules.

Mood Monitoring

A total of 66 apps inquired about users' current mood, of which 20 (30%) apps queried about reasons for low or depressed mood. In addition, 32 mental health and depression apps administered a screening questionnaire, particularly the depression module of the Patient Health Questionnaire (PHQ; PHQ-9: 20/32, 63%; PHQ-8: 2/32, 6%) alone or accompanied by another questionnaire, including Generalized Anxiety Disorder-7 (13 apps), Work and Social Adjustment Scale (3 apps), and Depression, Anxiety, and Stress Scale-21 (1 app). Moreover, 6% (2/32) of apps administered only Depression, Anxiety, and Stress Scale-21, and 25% (8/32) of apps administered a nonvalidated questionnaire.

Suicide Risk Management

Just over a third (34/98, 35%) of apps acknowledged suicide risk associated with depression by listing crisis management resources or actively inquiring about suicide risk. A total of 3% (2/65) of mental health apps and 8% (1/13) of depression apps directly asked users about suicidal thoughts. Other apps, including 3% (2/65) of mental health and 31% (4/13) of depression apps passively assessed suicide risk through question 9 of the PHQ-9 ("thoughts that you would be better off dead or of hurting yourself in some way"). Although 4 apps responded to a positive answer to question 9 by offering access to crisis helplines, 2 apps did not actively respond to PHQ-9 results, instead offering a list of suicide prevention resources. A total of 6% (6/65) of mental health apps were conversational agents that responded to users' suicidal thoughts by offering access to crisis helplines directly through the app. Finally, 10% (2/20) of well-being apps, 20% (13/65) of mental health apps, and 31% (4/13) of depression apps presented a list of crisis helpline numbers without assessing suicide risk.

Homework Assignment

A total of 20% (20/98) of apps offered users homework activities. In total, 70% (14/20) of apps assigned varied homework tasks, such as mood monitoring, reframing thoughts, BA, and relaxation exercises, whereas 30% (6/20) of apps consisted of a homework template to be completed between therapist visits.

Therapeutic Alliance

Overall, 46% (46/98) of apps included one or more engagement strategies comprising push notifications or reminders to access app features (31/46, 67%), offering feedback or encouraging messages (32/46, 70%) and/or awarding badges or other virtual rewards for completed activities (18/46, 39%). One conversational agent [66] was able to remember past conversations to further personalize the interaction.

Other Features

A total of 48% (47/98) of apps, including 45% (9/20) of well-being apps, 49% (32/65) of mental health apps, and 46% (6/13) of depression apps, also included one or more non–CBT-based features, particularly unstructured writing in the form of journals, gratitude, or positive thoughts; web-based games; and forums to connect to peers. A variety of other non–CBT-based functionalities were present in 1 or 2 assessed apps. These included inspiring quotes, visualization techniques, personality tests, access to self-help books, food records, management strategies for anxiety and panic attacks, safety plan templates, medication reminders, acceptance and self-compassion modules, music tracks, videos, physical exercise routines, and jokes.

Technical Aspects and Quality Assurance

Overview

In general, the assessed apps were easy to use and provided a logical and simple layout to navigate. Most apps worked as intended, and only 6% (6/98) apps either crashed while in use or included nonfunctional features. None of the assessed apps included advertisements. A total of 46% (45/98) of apps allowed users to share data with health care providers, members of their support network, or other users of the app, using email or data syncing. Table 4 provides a summary of technical aspects and quality assurance of the apps.

XSL•FO
RenderX

Table 4. Technical features and quality assurance of included apps (N=98).

Feature	App category, n (%)			Total (N=98), n (%)
	Well-being (n=20)	Mental health (n=65)	Depression (n=13)	
App credibility				
App content referenced or signed by the author	0 (0)	29 (45)	9 (69)	38 (39)
App includes a disclaimer that app information does not replace health care provider's advice	5 (25)	41 (63)	9 (69)	55 (56)
App development team mentioned in app content				
Government agency, academic institution, or nongovernment organization	2 (10)	4 (6)	0 (0)	6 (6)
Health care professional	1 (5)	36 (55)	11 (85)	48 (49)
Not declared	17 (85)	25 (38)	2 (15)	44 (45)
Data privacy				
Authentication required to access app	12 (60)	46 (71)	9 (69)	67 (68)
App with a privacy policy[a]	18 (90)	59 (91)	12 (92)	89 (91)
Presented before account creation	5 (6)	26 (29)	3 (25)	34 (35)
Explains how data are collected	17 (94)	58 (98)	12 (100)	87 (98)
Shares information with third-party providers	14 (78)	51 (86)	9 (75)	74 (83)
Shares contact details of Data Protection Officer	7 (35)	13 (22)	0 (0)	20 (22)
App allows users to share data	4 (20)	35 (59)	6 (50)	45 (51)

[a]Percentage calculation was based on the number of apps with privacy policy (well-being apps: 18; mental health apps: 59; depression apps: 12; total: 89).

App Credibility

App content was referenced by 58% (57/98) of apps, including 10% (2/20) of well-being apps, 68% (44/65) of mental health apps, and 85% (11/13) of depression apps. Overall, just over half (54/98, 55%) of the apps were developed by academic institutions or included psychiatrists or psychologists in the development team. Disclaimers were present in 55 apps.

A total of 18 publications in peer-reviewed journals were available for 17% (17/98) of apps, including 25% (16/65) of mental health apps and 8% (1/13) of depression apps. A total of 33% (6/18) of papers included randomized controlled trials [67-72] reporting that apps or some specific app functions were effective in improving mood, anxiety, stress, or general well-being compared with active or waitlist controls. Other publications included pilot studies (3/18, 17%) [73-75], secondary data analysis (1/18, 6%) [76], and app usage data analyses (8/18, 44%) [77-84] to assess user engagement and/or app effectiveness in decreasing users' symptoms.

Data Privacy and Security

Two-thirds (67/98, 68%) of the apps required authentication to access the app in the form of a password or less often a 4- to 6-digit numerical PIN. In addition, 49% (48/98) of apps allowed users to customize the privacy settings in the app.

Most apps (90/98, 92%) offered a privacy policy accessible from the app itself or through a link from the app store. The privacy policy was presented to users before account creation in only one-third of the apps (34/98, 35%). Most privacy policies (88/90, 98%) explained how users' personal data were collected, and more than three-fourth of which (74/90, 82%) stated that they shared data with third parties, often service providers. Nonetheless, only a few apps mentioned the names of the service providers or explicitly stated the type of data shared with them. A total of 2 apps stated that they shared data with advertising companies. Just over half of the privacy policies (47/90, 52%) addressed the requirements of the General Data Protection Regulation mentioning the users' right to have their data corrected or deleted, whereas only 24% (22/90) privacy policies included the contact details of a data protection officer.

Discussion

Principal Findings

Our systematic assessment of 98 Android (including 5 apps downloaded more than 1,000,000 times) and iOS self-guided CBT apps revealed a heterogeneous group offering a range of evidence-based [53,54] and non–evidence-based CBT techniques.

Only 4 mental health apps offered all five evidence-based CBT techniques. Depression apps consistently offered three to four techniques, whereas most well-being and mental health apps offered two evidence-based techniques, suggesting that only a few apps currently offer comprehensive, self-guided CBT programs that may benefit users who are unable to access face-to-face psychotherapy. Cognitive restructuring was the

XSL•FO
RenderX

most common technique across all app categories. Psychoeducation and BA were offered by most depression apps and approximately half of mental health apps; however, they were seldomly included in well-being apps. In addition, well-being and mental health apps frequently offered relaxation and mindfulness, whereas less than half of depression apps did so.

Our study's broad inclusion criteria aimed to resemble the options offered by app marketplaces when users search for a self-help app. Furthermore, classifying the apps into three distinct groups revealed that depression apps consistently offered more comprehensive programs, including at least three evidence-based techniques, whereas mental health and well-being apps were substantially less adherent. This finding differs from previous studies evaluating mental health apps [52,57,85,86] and, more specifically, CBT for depression [47,48] apps that consistently reported low adherence to evidence-based techniques. Huguet et al [47] evaluated 12 CBT and BA apps for depression available in Canada and reported a median adherence to core CBT criteria of 15%, with 2 apps adhering to 75% of the criteria. Stawarz et al [48] analyzed 31 CBT apps and reported that most apps included cognitive restructuring and offered one or two CBT techniques.

Only one-third of the assessed apps offered resources to address suicide risk, whereas only a minority actively assessed users' suicide risk. App developers appear to disregard suicide risk when designing well-being and mental health apps, a worrying trend considering that most people dying by suicide are affected by a mental disorder [87,88].

Although most apps included a privacy policy mentioning how users' personal data were used and shared, only a few provided sufficient details on the type of data shared or the companies with which the data were shared. Previous studies have shown that app developers often share user data with third parties, including Google and Facebook advertising [45,89], even when this is not stated in the privacy policy. Despite increasing concerns with regard to data privacy and security, apps still present considerable data management shortcomings, such as allowing third-party services to install pieces of code in the app to secretly access user data [90], as data sharing is a source of revenue, particularly for free apps. Notwithstanding the repeated calls for improvement, app development and publication processes remain unclear, allowing for loose interpretation or disregard of regulations. Further steps still need to be taken to ensure that user data are not misused, particularly highly sensitive data such as mental health–related information.

None of the assessed apps offered personalized content beyond including usernames or sex-appropriate pronouns when offering feedback or during dialogs with conversational agents. As a proxy to protect users' privacy, apps did not ask personal questions beyond name, sex, or age, nor did they inquire about relevant medical history. Other engagement features, such as push notifications, reminders, and gamification, were found in less than half of the included apps. Although the role of the *bond* between users and interventions in self-guided digital interventions remains unclear, it appears that tailoring content, personalization, and interactive features such as reminders, positive feedback, and supporting social interactions with other app users improve engagement and adherence [55].

One of the advantages of digital technologies is the possibility of continuous updates and improvements on the basis of contextual challenges or new clinical guidelines. However, as only 17% (17/98) of the included apps included information on the impact of COVID-19 and containment measures on mental health [17], it can be assumed that digital mental health providers do not fully leverage available technological opportunities.

Publications in peer-reviewed journals were available for 17 apps, of which only 6 (35%) apps were evaluated using a randomized clinical trial, confirming the lack of evidence supporting the use of publicly available apps [91,92]. When evidence is available, clinical trials have consistently shown that apps improve mood in users affected by depression [67,68] during and shortly after the completion of the trial, even if high dropout rates limit the validity of the results [44].

This study, using broad inclusion criteria, showed a great variety of CBT-based apps, from simple journaling apps to structured modules mirroring traditional CBT. Although such extensive selection may positively affect a larger number of users, it might be cumbersome and frustrating for users that specifically require a more structured self-guided CBT program, potentially hampering access to self-help tools. Mental health apps follow a continuum from well-being and lifestyle apps for healthy users to apps offering self-guided psychotherapy as an adjuvant to ongoing traditional therapy or as the sole therapy for people with mental health disorders. As such, the current app store categories may not be suitable for mental health apps. We propose the creation of a new, encompassing mental health category to include all well-being and mental health specific apps with subcategories targeting different mental health disorders, therapy modalities, and user groups. In addition, a wider overhaul of health app development and publication processes is required.

Strengths and Limitations of the Study

Our study had several strengths. Our process for searching, retrieving, and assessing included apps is grounded in systematic review methodology. The app search was implemented using the search engine of a commercial app database that allowed for a geographically unrestricted app search. Comprehensive assessment criteria were developed using reputable CBT manuals and a framework developed by our center to ensure comprehensive results.

However, this study has several limitations. Our search strategy included only CBT-related terms, potentially missing other apps offering CBT-based therapy for users with low mood or depression. We included apps offering self-guided CBT interventions, given the ethical limitations of using a simulated app persona when interacting with a health care provider, potentially excluding from our analysis more comprehensive apps. Apps linked to a specific provider or requiring an access code were also excluded from our study. Furthermore, the app store search was limited to only four categories with the maximum probability of retrieving relevant apps; therefore, we

XSL•FO
RenderX

may have missed other apps outside these categories. Only apps in English were included in our study, potentially excluding relevant apps in other languages.

Conclusions

Self-guided CBT-based apps available in app marketplaces offer a wide range of interventions; however, only approximately one-third, particularly depression apps, included comprehensive CBT programs. App developers' access, use, and sharing of user data are unclear, raising concerns about the privacy and security of user data, and highlighting severe shortcomings in the governance of the health app market. Only a few apps offered suicide risk management resources or information on the current COVID-19 pandemic. The classification of mental health apps may benefit from the creation of a new mental health app category, including all well-being and specific disorder apps. To fulfill their potential, it is essential that self-guided CBT-based apps adhere to evidence-based clinical guidelines, be patient-centered, and offer enhanced and transparent data security measures.

Acknowledgments

LM gratefully acknowledges the National Technological University Research Scholarship (Lee Kong Chian School of Medicine) support for her PhD studentship, which enabled this work. JC's post at Imperial College London is supported by the National Institute for Health Research Northwest London Applied Research Collaboration. FVW and JC are supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors' Contributions

LM had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. LM, KG, and JC provided the concept and design of the study. LM acquired, analyzed and interpreted the data; performed statistical analysis; and provided administrative, technical and material support. The manuscript was drafted by LM and ACS. All authors contributed for critical revision of the manuscript for important intellectual content. JC obtained the funding. KG and JC were responsible for supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Assessment criteria for cognitive behavioral therapy–related features.
[DOCX File , 35 KB - jmir_v23i7e27619_app1.docx]

Multimedia Appendix 2
Characteristics of included apps.
[DOCX File , 54 KB - jmir_v23i7e27619_app2.docx]

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Stigma and stigmatization within and beyond psychiatry. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf;jsessionid=BB5A000A2E96A35EDC53EF0465D1F625?sequence=1 [accessed 2018-12-13]
3. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
4. Luitel NP, Jordans MJ, Kohrt BA, Rathod SD, Komproe IH. Treatment gap and barriers for mental health care: a cross-sectional community survey in Nepal. PLoS One 2017;12(8):e0183223 [FREE Full text] [doi: 10.1371/journal.pone.0183223] [Medline: 28817734]
5. Subramaniam M, Abdin E, Vaingankar JA, Shafie S, Chua HC, Tan WM, et al. Minding the treatment gap: results of the Singapore Mental Health Study. Soc Psychiatry Psychiatr Epidemiol 2020 Nov;55(11):1415-1424 [FREE Full text] [doi: 10.1007/s00127-019-01748-0] [Medline: 31317246]
6. Shi W, Shen Z, Wang S, Hall BJ. Barriers to professional mental health help-seeking among chinese adults: a systematic review. Front Psychiatry 2020;11:442 [FREE Full text] [doi: 10.3389/fpsyt.2020.00442] [Medline: 32508688]

XSL•FO
RenderX

7. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
8. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
9. Prins M, Meadows G, Bobevski I, Graham A, Verhaak P, van der Meer K, et al. Perceived need for mental health care and barriers to care in the Netherlands and Australia. Soc Psychiatry Psychiatr Epidemiol 2011 Oct;46(10):1033-1044 [FREE Full text] [doi: 10.1007/s00127-010-0266-3] [Medline: 20686887]
10. Collins KA, Westra HA, Dozois DJ, Burns DD. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. Clin Psychol Rev 2004 Sep;24(5):583-616. [doi: 10.1016/j.cpr.2004.06.001] [Medline: 15325746]
11. Raingruber B. Gaps in service in the recognition and treatment of depression and suicidal ideation within a four-county area. Perspect Psychiatr Care 2003;39(4):151-162. [doi: 10.1111/j.1744-6163.2003.00151.x] [Medline: 14737821]
12. Wang J. Perceived barriers to mental health service use among individuals with mental disorders in the Canadian general population. Med Care 2006 Feb;44(2):192-195. [doi: 10.1097/01.mlr.0000196954.67658.95] [Medline: 16434920]
13. Psychiatrists and Nurses (Per 100 000 Population). World Health Organization. 2019. URL: https://apps.who.int/gho/data/node.main.MHHR?lang=en [accessed 2020-08-11]
14. Navarro-Pérez JJ, Mestre M. Challenges and barriers in mental healthcare systems and their impact on the family: a systematic integrative review. Health Soc Care Community 2020 Sep;28(5):1366-1379. [doi: 10.1111/hsc.12968] [Medline: 32115797]
15. Mental health services in Australia. Australia: Australian Institute of Health and Welfare URL: https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/mental-health-workforce [accessed 2020-08-26]
16. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
17. Campion J, Javed A, Sartorius N, Marmot M. Addressing the public mental health challenge of COVID-19. Lancet Psychiatry 2020 Aug;7(8):657-659 [FREE Full text] [doi: 10.1016/S2215-0366(20)30240-6] [Medline: 32531299]
18. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. Lancet Psychiatry 2020 Oct;7(10):883-892 [FREE Full text] [doi: 10.1016/S2215-0366(20)30308-4] [Medline: 32707037]
19. David D, Cristea I, Hofmann SG. Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
20. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
21. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
22. Santoft F, Axelsson E, Öst L, Hedman-Lagerlöf M, Fust J, Hedman-Lagerlöf E. Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. Psychol Med 2019 Jun;49(8):1266-1274. [doi: 10.1017/S0033291718004208] [Medline: 30688184]
23. Wenzel A, Brown G, Karlin B. Cognitive behavioral therapy for depression in veterans and military servicemembers: Therapist manual. Washington, DC: U.S. Department of Veterans Affairs; 2011. URL: https://www.mirecc.va.gov/docs/cbt-d_manual_depression.pdf [accessed 2020-06-29]
24. Mennin DS, Ellard KK, Fresco DM, Gross JJ. United we stand: emphasizing commonalities across cognitive-behavioral therapies. Behav Ther 2013 Jun;44(2):234-248. [doi: 10.1016/j.beth.2013.02.004] [Medline: 23611074]
25. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
26. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:46-54 [FREE Full text] [Medline: 15231555]
27. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [Medline: 22517917]
28. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]

XSL•FO
RenderX

29. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User Acceptance of Computerized Cognitive Behavioral Therapy for Depression: Systematic Review. J Med Internet Res 2017 Sep 13;19(9):e309. [doi: 10.2196/jmir.7662] [Medline: 28903893]
30. Wells MJ, Owen JJ, McCray LW, Bishop LB, Eells TD, Brown GK, et al. Computer-Assisted Cognitive-Behavior Therapy for Depression in Primary Care: Systematic Review and Meta-Analysis. Prim Care Companion CNS Disord 2018 Mar 01;20(2):17r02196 [FREE Full text] [doi: 10.4088/PCC.17r02196] [Medline: 29570963]
31. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
32. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
33. Topol E. Digital medicine: empowering both patients and clinicians. Lancet 2016 Aug 20;388(10046):740-741. [doi: 10.1016/S0140-6736(16)31355-1] [Medline: 27560260]
34. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
35. Inkster B, O'Brien R, Selby E, Joshi S, Subramanian V, Kadaba M, et al. Digital Health Management During and Beyond the COVID-19 Pandemic: Opportunities, Barriers, and Recommendations. JMIR Ment Health 2020 Jul 06;7(7):e19246 [FREE Full text] [doi: 10.2196/19246] [Medline: 32484783]
36. Aitken M, Clancy B, Nass D. The growing value of digital health - Evidence and impact on human health and the healthcare system. USA: IQVIA Institute for Human Data Science. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health#:~:text=The%20use%20of%20such%20Digital,estimated%20%247%20billion%20per%20year [accessed 2018-07-11]
37. Torous J, Roberts LW. Needed Innovation in Digital Health and Smartphone Applications for Mental Health: Transparency and Trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
38. Newzoo. Penetration rate of smartphones in selected countries 2020 Graph. Statista 2020. URL: https://www.statista.com/statistics/539395/smartphone-penetration-worldwide-by-country/ [accessed 2021-04-05]
39. Rebello TJ, Marques A, Gureje O, Pike KM. Innovative strategies for closing the mental health treatment gap globally. Curr Opin Psychiatry 2014 Jul;27(4):308-314. [doi: 10.1097/YCO.0000000000000068] [Medline: 24840160]
40. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
41. Vo V, Auroy L, Sarradon-Eck A. Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
42. Rathbone AL, Clarry L, Prescott J. Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
43. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile Apps for Mental Health Issues: Meta-Review of Meta-Analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
44. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
45. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
46. Bates DW, Landman A, Levine DM. Health Apps and Health Policy: What Is Needed? JAMA 2018 Nov 20;320(19):1975-1976. [doi: 10.1001/jama.2018.14378] [Medline: 30326025]
47. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
48. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User Experience of Cognitive Behavioral Therapy Apps for Depression: An Analysis of App Functionality and User Reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
49. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]

XSL•FO
RenderX

50. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision Support and Alerts of Apps for Self-management of Blood Glucose for Type 2 Diabetes. JAMA 2019 Dec 16;321(15):1530-1532. [doi: 10.1001/jama.2019.1644] [Medline: 30990543]
51. van Galen LS, Xu X, Koh MJA, Thng S, Car J. Eczema apps conformance with clinical guidelines: a systematic assessment of functions, tools and content. Br J Dermatol 2020 Feb;182(2):444-453. [doi: 10.1111/bjd.18152] [Medline: 31179535]
52. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. BMC Med 2019 Dec 19;17(1):231 [FREE Full text] [doi: 10.1186/s12916-019-1461-z] [Medline: 31852455]
53. Beck J. Cognitive behavior therapy: Basics and beyond, 2nd ed. New York, NY. US: Guilford Press; 2011.
54. Roth A, Pilling S. The competences required to deliver effective cognitive and behavioural therapy for people with depression and with anxiety disorders. Improving Access to Psychological Therapies (IAPT) Programme, Department of Health. 2007. URL: https://www.uea.ac.uk/documents/20142/605327/norwich-medical-school-the-competences-required-to-deliver-effective-cognitive-and- behavioural-therapy-for-people-with-depression-and-with-anxiety-disorders.pdf/feaa5fd3-fee7-c597-c085-d2c98329edac?t=1576402736010 [accessed 2020-02-10]
55. Tremain H, McEnery C, Fletcher K, Murray G. The Therapeutic Alliance in Digital Mental Health Interventions for Serious Mental Illnesses: Narrative Review. JMIR Ment Health 2020 Aug 07;7(8):e17204 [FREE Full text] [doi: 10.2196/17204] [Medline: 32763881]
56. 42matters. URL: https://42matters.com/ [accessed 2020-02-19]
57. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9. [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
58. eGuru Depression. URL: https://play.google.com/store/apps/details?id=com.eguru4life.egurudepression&hl=en_SG [accessed 2020-07-27]
59. eGuru Mood Diary. URL: https://play.google.com/store/apps/details?id=com.eguru4life.egurumooddiary&hl=en_SG [accessed 2020-03-20]
60. eGuru Thought Diary. URL: https://play.google.com/store/apps/details?id=com.eguru4life.eguruthoughtdiary&hl=en_SG [accessed 2020-07-27]
61. Moodpath - Depression & Anxiety Test. URL: https://play.google.com/store/apps/details?id=de.moodpath.android&hl=en_SG [accessed 2020-03-11]
62. Sanvello for Anxiety, Depression & Stress. URL: https://play.google.com/store/apps/details?id=com.pacificalabs.pacifica [accessed 2020-03-12]
63. Wysa: stress & anxiety therapy chatbot. URL: https://play.google.com/store/apps/details?id=bot.touchkin&hl=en_SG [accessed 2020-05-15]
64. Youper - Anxiety & Depression. URL: https://play.google.com/store/apps/details?id=br.com.youper&hl=en_SG [accessed 2020-05-15]
65. Reflectly - Journal / Diary. URL: https://play.google.com/store/apps/details?id=com.reflectlyApp [accessed 2020-03-16]
66. Happify. Happify. URL: https://www.happify.com/ [accessed 2020-05-15]
67. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
68. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached Mobile App Platform for the Treatment of Depression and Anxiety Among Primary Care Patients: A Randomized Clinical Trial. JAMA Psychiatry 2020 Sep 01;77(9):906-914. [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
69. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot Randomized Trial of a Self-Help Behavioral Activation Mobile App for Utilization in Primary Care. Behav Ther 2019 Jul;50(4):817-827. [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
70. Hunter JF, Olah MS, Williams AL, Parks AC, Pressman SD. Effect of Brief Biofeedback via a Smartphone App on Stress Recovery: Randomized Experimental Study. JMIR Serious Games 2019 Nov 26;7(4):e15974 [FREE Full text] [doi: 10.2196/15974] [Medline: 31769761]
71. Moberg C, Niles A, Beermann D. Guided Self-Help Works: Randomized Waitlist Controlled Trial of Pacifica, a Mobile App Integrating Cognitive Behavioral Therapy and Mindfulness for Stress, Anxiety, and Depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
72. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
73. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and Pilot Evaluation of Smartphone-Delivered Cognitive Behavior Therapy Strategies for Mood- and Anxiety-Related Problems: MoodMission. Cognitive and Behavioral Practice 2018 Nov;25(4):496-514. [doi: 10.1016/j.cbpra.2018.07.002]

XSL•FO
RenderX

74. Mehrotra S, Sudhir P, Rao G, Thirthalli J, Srikanth TK. Development and Pilot Testing of an Internet-Based Self-Help Intervention for Depression for Indian Users. Behav Sci (Basel) 2018 Mar 22;8(4):36 [FREE Full text] [doi: 10.3390/bs8040036] [Medline: 29565278]
75. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: An Eclectic, Skills-Based App Suite for the Treatment of Depression and Anxiety. J Med Internet Res 2017 Jan 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
76. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically Meaningful Use of Mental Health Apps and its Effects on Depression: Mixed Methods Study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
77. Kinderman P, Hagan P, King S, Bowman J, Chahal J, Gan L, et al. The feasibility and effectiveness of Catch It, an innovative CBT smartphone app. BJPsych Open 2016 May;2(3):204-209 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002436] [Medline: 27703777]
78. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
79. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the "Big" Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data. J Med Internet Res 2016 Aug 31;18(8):e241 [FREE Full text] [doi: 10.2196/jmir.5725] [Medline: 27580524]
80. Kwasny MJ, Schueller SM, Lattie E, Gray EL, Mohr DC. Exploring the Use of Multiple Mental Health Apps Within a Platform: Secondary Analysis of the IntelliCare Field Trial. JMIR Ment Health 2019 Mar 21;6(3):e11572 [FREE Full text] [doi: 10.2196/11572] [Medline: 30896433]
81. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and Usage of IntelliCare: A Publicly Available Suite of Mental Health and Well-Being Apps. Internet Interv 2016 May;4(2):152-158 [FREE Full text] [doi: 10.1016/j.invent.2016.06.003] [Medline: 27398319]
82. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The Effects of a Digital Well-Being Intervention on Patients With Chronic Conditions: Observational Study. J Med Internet Res 2020 Jan 10;22(1):e16211 [FREE Full text] [doi: 10.2196/16211] [Medline: 31922491]
83. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Aust Psychol 2019 Mar 13;54(4):245-260. [doi: 10.1111/ap.12383]
84. Scherr S, Goering M. Is a Self-Monitoring App for Depression a Good Place for Additional Mental Health Information? Ecological Momentary Assessment of Mental Help Information Seeking among Smartphone Users. Health Communication 2019 Apr 26:1-9. [doi: 10.1080/10410236.2019.1606135]
85. Huckvale K, Nicholas J, Torous J, Larsen ME. Smartphone apps for the treatment of mental health conditions: status and considerations. Curr Opin Psychol 2020 Dec;36:65-70 [FREE Full text] [doi: 10.1016/j.copsyc.2020.04.008] [Medline: 32553848]
86. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
87. World Health Organization. Preventing suicide: A global imperative. Geneva: World Health Organization; 2014. URL: https://www.who.int/mental_health/suicide-prevention/world_report_2014/en/ [accessed 2019-01-03]
88. Hawton K, Casañas I Comabella C, Haw C, Saunders K. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord 2013 May;147(1-3):17-28. [doi: 10.1016/j.jad.2013.01.004] [Medline: 23411024]
89. Rosenfeld L, Torous J, Vahia IV. Data Security and Privacy in Apps for Dementia: An Analysis of Existing Privacy Policies. Am J Geriatr Psychiatry 2017 Aug;25(8):873-877. [doi: 10.1016/j.jagp.2017.04.009] [Medline: 28645535]
90. Razaghpanah A, Nithyanand R, Vallina-Rodriguez N. Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem. 2018 Presented at: Proceedings of the Network and Distributed Systems Security (NDSS) Symposium; 2018; San Diego, CA, USA p. 1-15. [doi: 10.14722/ndss.2018.23353]
91. Marshall JM, Dunstan DA, Bartik W. Apps With Maps-Anxiety and Depression Mobile Apps With Evidence-Based Frameworks: Systematic Search of Major App Stores. JMIR Ment Health 2020 Jun 24;7(6):e16525 [FREE Full text] [doi: 10.2196/16525] [Medline: 32579127]
92. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]

Abbreviations

BA: behavioral activation
CBT: cognitive behavioral therapy
iCBT: internet-based cognitive behavioral therapy
PHQ: Patient Health Questionnaire

XSL•FO
RenderX

Edited by C Basch; submitted 31.01.21; peer-reviewed by S Moritz, A Graham; comments to author 03.04.21; revised version received 13.04.21; accepted 23.04.21; published 30.07.21.

Please cite as:
Martinengo L, Stona AC, Griva K, Dazzan P, Pariante CM, von Wangenheim F, Car J
Self-guided Cognitive Behavioral Therapy Apps for Depression: Systematic Assessment of Features, Functionality, and Congruence With Evidence
J Med Internet Res 2021;23(7):e27619
URL: https://www.jmir.org/2021/7/e27619
doi:10.2196/27619
PMID:34328431

©Laura Martinengo, Anne-Claire Stona, Konstadina Griva, Paola Dazzan, Carmine Maria Pariante, Florian von Wangenheim, Josip Car. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Smart Diaper System Using Bluetooth and Smartphones to Automatically Detect Urination and Volume of Voiding: Prospective Observational Pilot Study in an Acute Care Hospital

Jae Ho Cho[1], MSc, MD; Jung-Yeon Choi[1], MD, PhD; Nak-Hyun Kim[1], MD, PhD; Yejee Lim[1], MD, PhD; Jung Hun Ohn[1], MD, PhD; Eun Sun Kim[1], MSc, MD; Jiwon Ryu[1], MSc, MD; Jangsun Kim[2], BS; Yiseob Kim[2], MSc; Sun-wook Kim[1], MSc, MD; Kwang-Il Kim[1], MD, PhD

[1]Department of Internal Medicine, Seoul National University Bundang Hospital, Gyeonggi-do, Republic of Korea

[2]Panoptics Corp, Gyeonggi-do, Republic of Korea

Corresponding Author:
Sun-wook Kim, MSc, MD
Department of Internal Medicine
Seoul National University Bundang Hospital
82 Gumi-ro, 173 Beon-gil
Bundang-gu, Seongnam
Gyeonggi-do, 13620
Republic of Korea
Phone: 82 31 787 7813
Fax: 82 31 787 4051
Email: apollo19@snu.ac.kr

Abstract

Background: Caregivers of patients who wear conventional diapers are required to check for voiding every hour because prolonged wearing of wet diapers causes health problems including diaper dermatitis and urinary tract infections. However, frequent checking is labor intensive and disturbs patients' and caregivers' sleep. Furthermore, assessing patients' urine output with diapers in an acute care setting is difficult. Recently, a smart diaper system with wetness detection technology was developed to solve these issues.

Objective: We aimed to evaluate the applicability of the smart diaper system for urinary detection, its accuracy in measuring voiding volume, and its effect on incontinence-associated dermatitis (IAD) occurrence in an acute care hospital.

Methods: This prospective, observational, single-arm pilot study was conducted at a single tertiary hospital. We recruited 35 participants aged ≥50 years who were wearing diapers due to incontinence between August and November 2020. When the smart diaper becomes wet, the smart diaper system notifies the caregiver to change the diaper and measures voiding volume automatically. Caregivers were instructed to record the weight of wet diapers on frequency volume charts (FVCs). We determined the voiding detection rate of the smart diaper system and compared the urine volume as automatically calculated by the smart diaper system with the volume recorded on FVCs. Agreement between the two measurements was estimated using a Bland-Altman plot. We also checked for the occurrence or aggravation of IAD and bed sores.

Results: A total of 30 participants completed the protocol and 390 episodes of urination were recorded. There were 108 records (27.7%) on both the FVCs and the smart diaper system, 258 (66.2%) on the FVCs alone, 18 (4.6%) on the smart diaper system alone, and 6 (1.5%) on the FVCs with sensing device lost. The detection rate of the smart diaper system was 32.8% (126/384). When analyzing records concurrently listed in both the FVCs and the smart diaper system, linear regression showed a strong correlation between the two measurements (R^2=0.88, P<.001). The Bland-Altman assessment showed good agreement between the two measurements, with a mean difference of –4.2 mL and 95% limits of agreement of –96.7 mL and 88.3 mL. New occurrence and aggravation of IAD and bed sores were not observed. Bed sores improved in one participant.

Conclusions: The smart diaper system showed acceptable accuracy for measuring urine volume and it could replace conventional FVCs in acute setting hospitals. Furthermore, the smart diaper system has the potential advantage of preventing IAD development and bed sore worsening. However, the detection rate of the smart diaper system was lower than expected. Detection rate polarization

XSL•FO
RenderX

among participants was observed, and improvements in the user interface and convenience are needed for older individuals who are unfamiliar with the smart diaper system.

(J Med Internet Res 2021;23(7):e29979) doi:10.2196/29979

KEYWORDS

smart diaper; urinary incontinence; enuresis; voided volume; diaper rash; smartphone; mobile phone; app; eHealth; mHealth; urine; medical device; sensor; prospective; pilot study; observational

Introduction

Background

Urinary incontinence (UI), defined as the involuntary or abnormal loss of urine, represents a major health care problem in older adults [1]. UI is associated with falls and fractures, sleep disorders, and urinary tract infections. In addition, it decreases activity of daily living (ADL) and confidence and increases immobilization and hospitalization [2]. The prevalence of UI in a community population ranges from 5% to 45% [1,3-5] and steadily increases with age [5,6]. Furthermore, more than half of nursing home residents have incontinence [7-9], which is almost twice as high as the prevalence of community-dwelling individuals. Various treatments for UI, such as behavioral intervention, medication, and surgery, are limited in efficacy and may pose a challenge for frail geriatric patients. As a consequence, for older adults who are unable to take care of themselves, the most feasible options are wearing absorbent products or indwelling urinary catheters [9,10]. A previous study showed that more than 85% and 77% of older patients who are in long-term care facilities prefer using diapers and urinary catheterization, respectively [11].

However, wearing a diaper poses a number of risk factors. Incontinence care practices, such as prescheduled routine diaper checks, particularly at night, could be uncomfortable for the patients and disturb their daily living or sleep [12]. Wearing a wet diaper for a long time can lead to other health issues including diaper dermatitis (known as incontinence-associated dermatitis [IAD]) and bacterial infection [13,14]. Furthermore, IAD and infection cause or worsen bed sores, especially in bedridden older patients [13]. Indwelling urinary catheters are an alternative for UI, are not associated with IAD, and allow for accurate measurement of urine output in the acute care setting. However, it can cause urinary tract infections, and prolonged catheterization is associated with immobilization, loss of ADL, longer hospital stays, and mortality [15-17]. The best approach to prevent these complications is to replace unnecessary urinary catheterization with diaper use and to both detect the soiled diaper and change it as soon as possible. However, changing absorbent products in a timely manner is labor intensive for caregivers and health care providers, as most older patients with diapers have dementia or cognitive impairment. Moreover, recognizing urination without patient notification is difficult. Regularly scheduled diaper checks are not effective because the incontinence episodes of patients do not occur at precise times. In acute care hospital settings, measuring the weight of a wet diaper to assess input/output volume status increases the burden on caregivers. The onerousness of changing diapers may lead to unnecessary urinary catheterization and adversely affect patients' morbidity and mortality.

Recently, the rapid development of Internet of Things (IoT) technology has changed intelligent health care systems. An increasing number of traditional health care services are being complemented or replaced with IoT [18]. Wireless technologies and miniaturized wearable sensing devices have contributed to several studies of smart diaper systems for incontinence care that detect moisture in absorbent products, inform caregivers to promptly change the wet diaper, and assess urinary output simultaneously [19-21]. Theoretically, the smart diaper system prevents the development of IAD or worsening of bed sores and provides a solution for urine output measurement without catheterization. Currently, some smart devices are being developed experimentally, but only a few clinical studies have been conducted on smart diapers. This study is the first to investigate the feasibility of a smart diaper system using Bluetooth and a smartphone to automatically detect urination and volume of voiding in an acute care hospital.

Objectives

In this study, we aimed to evaluate the feasibility and utility of the smart diaper as an alternative to the conventional diaper in a clinical setting. We determined the detection rate of participants' urination, accuracy of the urine output assessment, and occurrence of IAD. We also investigated the caregivers' and health care providers' experience and solicited suggestions.

Methods

Study Design and Population

This prospective, observational, single-arm pilot study was conducted at Seoul National University Bundang Hospital, a 1300-bed teaching hospital in Korea. From August to November 2020, participants were prospectively recruited at the hospitalist-run acute care unit and the geriatric center. Inclusion criteria were as follows: (1) patients aged ≥50 years, and (2) those who wore diapers due to incontinence. Participants who wore diapers temporarily at night or for part of the day, rather than 24 hours per day, were excluded. Demographic data, laboratory data, and previous medical history were retrieved from electronic medical records.

The participants received diaper pads for three days. They could extend the participation period up to two days, for a total of five days, if they wanted to continue using smart diapers on the third day of the study and had to remain hospitalized for medical reasons. The caregiver of each participant was provided with sets of the smart diaper system and trained daily by a researcher on how to use them. In response to the smart diaper system's

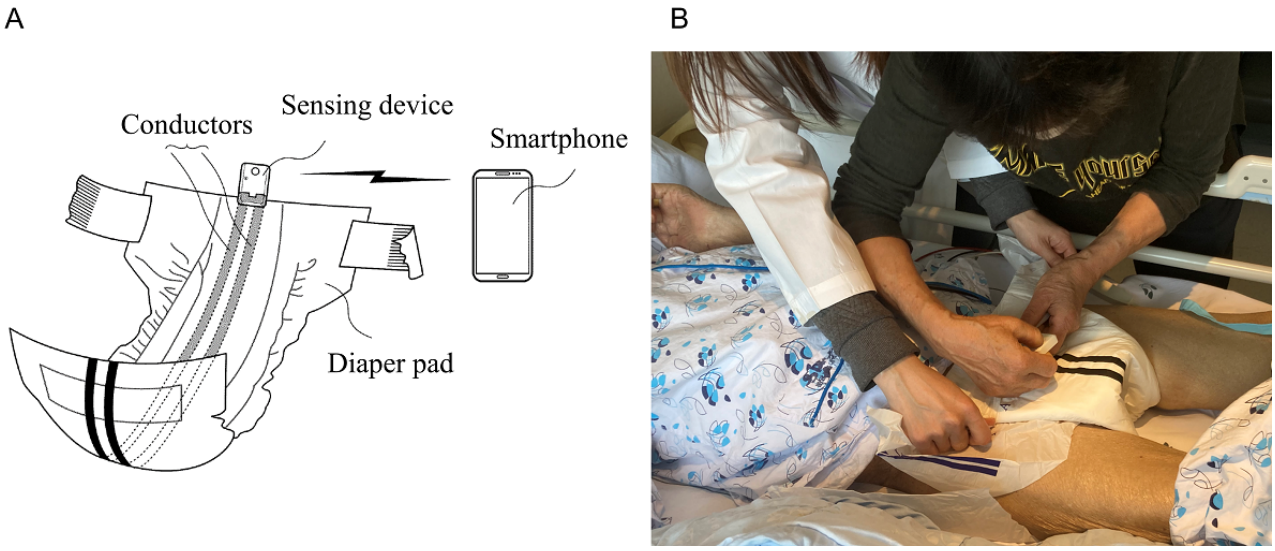
XSL•FO
RenderX

alarm, caregivers checked voiding, changed the diaper, weighed the wet diaper on a digital scale, and recorded the weight of the wet diaper on the frequency volume charts (FVCs). The researcher subtracted the weight of an unused diaper (110 g) from the weight of the wet diaper as recorded on the FVCs; this is considered the gold standard for determining urine volume. The manually measured urine output was compared with the urine output automatically detected by the smart diaper system to determine the accuracy of the smart diaper system. In addition, the researcher visited the participants once daily to check for the development or aggravation of IAD and record any evidence of IAD by taking photographs. Color, location, size, and character (scale, crust, discharge, etc) of IAD, as well as the presence of bed sores, were documented. Two investigators reviewed the participants' skin, and another investigator was consulted if any disagreement was found between the first two investigators. At the end of the trial, each caregiver completed a questionnaire on the user experience of the smart diaper system.

Materials

The smart diaper system consists of three parts: (1) a diaper embedded with conductors, (2) a sensing device, and (3) a smartphone (Figure 1). The inner surface of the diaper contains an absorbent liner for urine, and the outer surface has two lines of embedded conductors, which are printed at 1-cm intervals using carbon paste conductive ink. The two lines of conductors with regular spaces are connected to the sensing device at the front end of the diaper. The sensing device induces a minute electrical current, which a human cannot perceive, to the first conductor at regular time intervals. The current that passes through the second conductor is detected by the sensing device. When the diaper is wet, the current flow between the first and second conductors increases. As the amount of urine increased, the increment in amplitude of induced voltage increased. Therefore, the sensing device could detect subject's voiding and measure the amount of urine produced. The smartphone received a regular signal from the sensing device via Bluetooth. If the amount of urine exceeded a pre-established threshold, a dedicated app installed on the smartphone informed the caregiver about the patient's urination records. In this study, the minimum and maximum values of detectable urine volume were set to 50 mL and 500 mL, respectively. The diaper embedded with conductors obtained product certification as a personal hygiene item from the Korea Apparel Testing and Research Institute. There is a Korean patent pending for the whole system (10-2020-0093675).

Figure 1. The smart diaper system. (A) The design of the system consists of a diaper embedded with two lines of conductors, a sensing device, and a smartphone. (B) A caregiver connects the sensing device to a smart diaper under a researcher's guidance.



Ethics Statement

All procedures performed in this study involving human participants were in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study protocol was reviewed and approved by the Institutional Review Board of Seoul National University Bundang Hospital (B-2007/627-307). A researcher explained the study to the participants and their legal representatives and obtained written informed consent.

Statistical Analysis

Demographic data and baseline characteristics of the participants were analyzed using descriptive statistics. Data are presented as mean (SD) for continuous variables or counts with percentages for categorical variables. Linear regression analysis was used to evaluate the association between the manually weighed urine volume and the urine volume calculated by the smart diaper system. The correlation was described by the R^2 and plotted in a scatter plot. The agreement between the two measurements was analyzed by the Bland-Altman method [22,23]. The mean difference (bias) and upper and lower limits of agreement, defined as the mean difference ± 1.96 SD of differences, were calculated and constructed graphically with a Bland-Altman plot. A P value of <.05 was considered statistically significant, and all analyses were two-tailed. We logged and analyzed the data using SPSS (version 25.0; IBM Corp) and MedCalc (MedCalc Software). All data from the participants were deidentified and analyzed anonymously.

XSL•FO
RenderX

Calculation of a formal sample size was not performed in this study because this was a pilot study and estimable data from previous studies were unknown. However, a general rule of thumb is to include ≥30 patients to estimate a parameter [24]. Therefore, we recruited 35 patients, assuming approximately a 15% dropout rate.

Results

Baseline Characteristics

Of 35 participants, urinary catheterization was performed in three and two withdrew consent before data collection. Accordingly, data from 30 participants was used for the analysis (Figure 2). The demographic characteristics of the participants are described in Table 1. The mean age was 80.8 years, and 9 (30%) participants were men. Table 1 shows the commonly recorded diagnoses, including hypertension (23/30, 77%), diabetes (17/30, 57%), dementia (6/30, 20%), and Parkinsonism (2/30, 7%); in addition, 10 participants (33%) had a history of stroke. Serum total cholesterol, total protein, and albumin levels were lower than the normal range, suggesting risk of malnutrition.

Figure 2. Flow diagram of study participants.

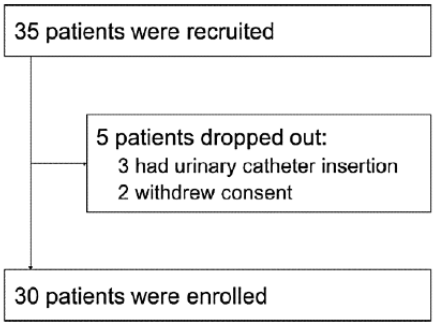


Table 1. Baseline characteristics of the study population.

Variables	Population		
	All (N=30)	Male (n=9)	Female (n=21)
Demographic data, mean (SD)			
Age (years)	80.8 (7.6)	82.1 (7.9)	80.2 (7.6)
Height (cm)	157.0 (7.8)	164.3 (4.7)	153.9 (6.8)
Weight (kg)	50.4 (8.1)	53.1 (7.9)	49.2 (8.0)
BMI (kg/m^2)	20.4 (2.9)	19.7 (3.1)	20.8 (2.9)
Laboratory data, mean (SD)			
White blood cell (×10^3/μL)	9.0 (4.0)	10.1 (4.4)	8.6 (3.9)
Hemoglobin (g/dL)	9.9 (1.7)	10.1 (2.1)	9.8 (1.6)
Platelet (×10^3/μL)	188.9 (93.7)	171.1 (111.8)	196.6 (86.8)
Blood urea nitrogen (mg/dL)	20.0 (9.6)	18.1 (7.2)	20.8 (10.5)
Creatinine (mg/dL)	0.9 (0.5)	1.0 (0.5)	0.9 (0.5)
Total cholesterol (mg/dL)	119.1 (38.7)	110 (48.6)	123.0 (34.2)
Protein (g/dL)	5.9 (1.0)	6.0 (1.3)	5.9 (0.8)
Albumin (g/dL)	2.9 (0.6)	2.7 (0.7)	3.0 (0.6)
Previous medical history, n (%)			
Hypertension	23 (77)	6 (67)	17 (81)
Diabetes	17 (57)	6 (67)	11 (52)
History of stroke	10 (33)	3 (33)	7 (33)
Dementia	6 (20)	2 (22)	4 (19)
Parkinsonism	2 (7)	1 (11)	1 (5)
Malignancy	11 (37)	4 (44)	7 (33)

XSL•FO
RenderX

Characteristics of Urination Records

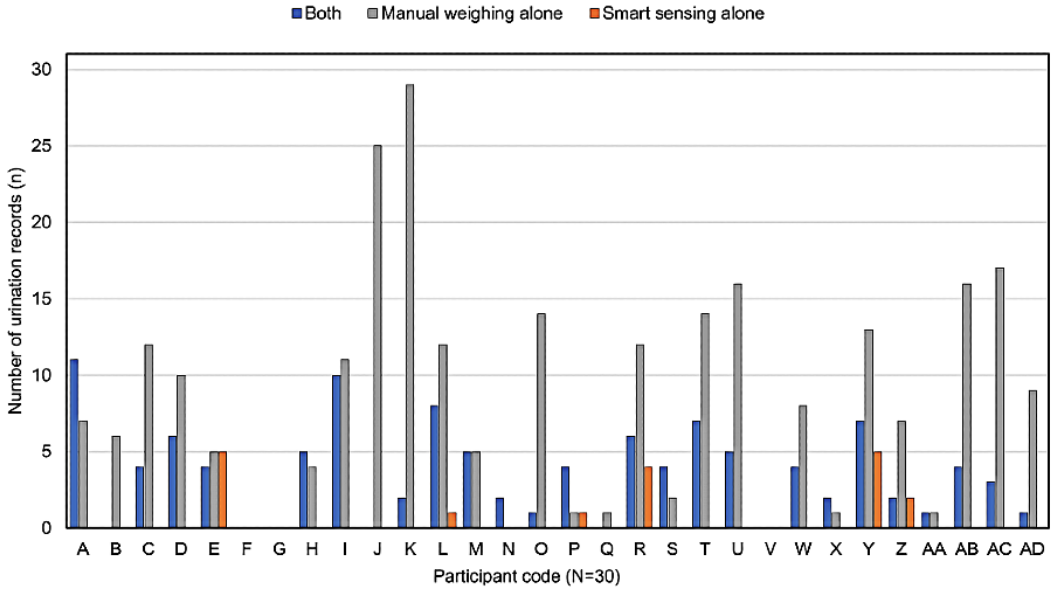
In total, 401 urination records found on the FVCs or the smart diaper system were collected, and 11 records reported 50 mL or less of urine. Since we had set the smart diaper system to not detect urine volumes less than 50 mL, we excluded those 11 records. Finally, 390 urination records were available for analysis. There was an average of 13.0 urination records per participant during the study period and an average of 3.7 urination records per person per day. The classification of 390 urination records is shown in Table 2. There were 108 paired records (27.7%) on the FVCs and the smart diaper system concurrently, 258 (66.2%) on the FVCs alone, 18 (4.6%) on the smart sensing system alone, and 6 (1.5%) on the FVCs with the sensing device lost so the smart diaper system could not detect urination. The detection rate of the smart diaper system was 32.8% (126/384), excluding 6 records where the sensing device was lost. The detailed urination records for each participant are shown in Figure 3. It should be noted that two participants (anonymous patient codes F and G) presented no record at all, and one participant (V) presented only urination records in which the urine volume measured by manual weighing was 50 mL or less.

Table 2. Classification of a total of 390 urination records collected from 30 participants[a].

Variables	Value (N=390), n (%)
Urination records, n (%)	
Both manual weighing and smart sensing records	108 (27.7)
Only manual weighing record	258 (66.2)
Only smart sensing record	18 (4.6)
Sensing device lost	6 (1.5)

[a]No data was available from 3 participants.

Figure 3. Histogram of distribution of urination records for each participant. There were no records available for 3 participants (F, G, and V).

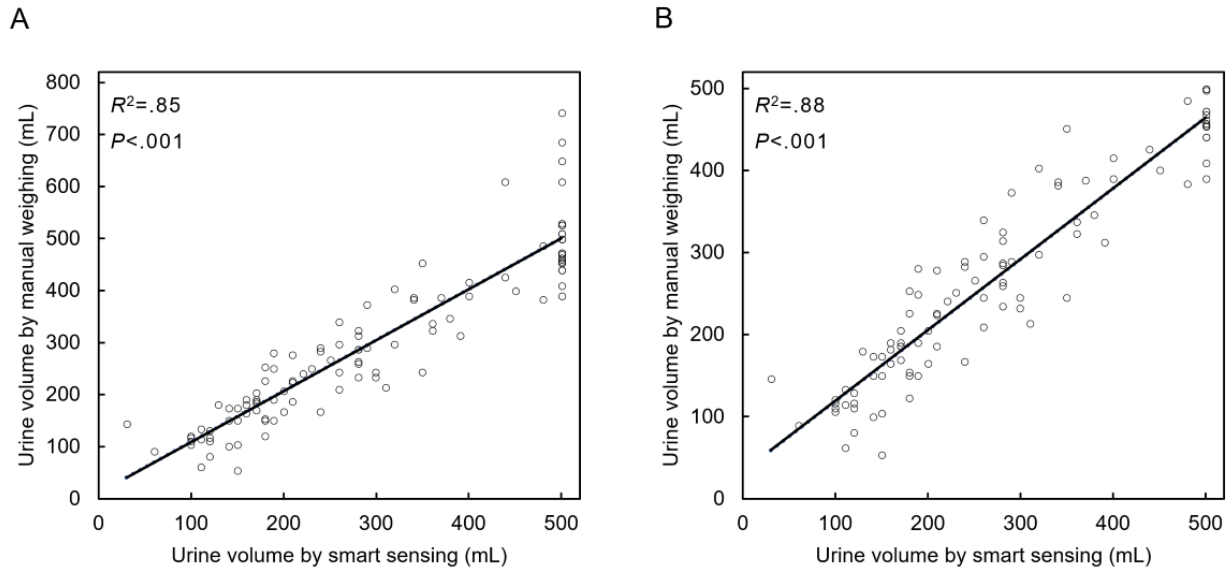


Correlation Between the Two Methods

To verify the accuracy in urine volume measurement of the smart diaper system, we evaluated the relationship and agreement between the two measurements: weighing wet diapers manually on a digital scale and automatic detection by smart sensing. In addition to the analysis of the 108 paired records, subgroup analysis was performed excluding 10 paired records containing a manually measured volume of voiding over 500 mL (ie, greater than the maximum detectable value for the smart diaper system). A linear regression showed a strong correlation between the two measurements for the 108 records (R^2=0.85, P<.001) and the 98 subgroup records (R^2=0.88, P<.001), respectively (Figure 4).

XSL•FO
RenderX

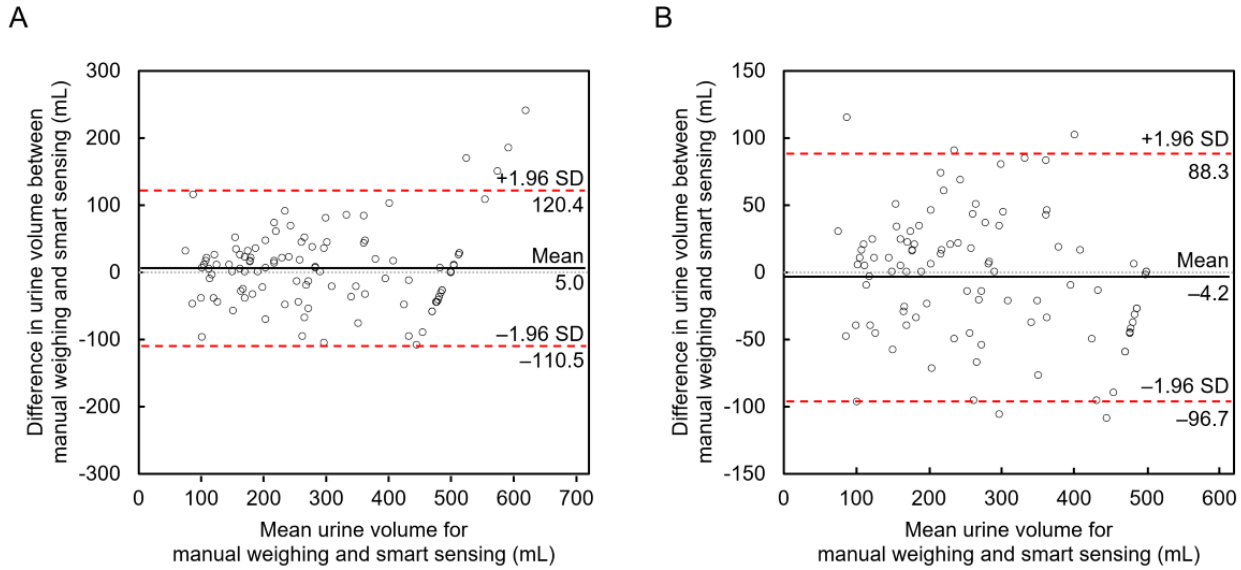
Figure 4. Scatter plot of urine volume measured by manual weighing and the smart diaper system. A significant linear regression relationship was found between the two measurements in the 108 urination records (A) and the 98 records (B).



Agreement Between the Two Methods

Bland-Altman analysis was performed for agreement between the two measurement techniques (Figure 5). In the analysis of the 108 records, the mean difference between the two measurements was 5.0 mL, and the lower and upper limits of agreement were –110.5 mL and 120.4 mL, respectively. In the analysis of the 98 subgroup records, the mean difference between the two measurements was –4.2 mL, and the lower and upper limits of agreement were –96.7 mL and 88.3 mL, respectively. Based on the analysis of the 98 subgroup records, the 95% limits of agreement of all paired data were deemed clinically interchangeable. Meanwhile, the average (SD) of the absolute values (|x|) of the difference between the two measurements was 42.6 mL (SD 40.9 mL) for the 108 records and 37.2 mL (SD 29.0 mL) for the 98 subgroup records, respectively.

Figure 5. Bland-Altman plot of urine volume measured by manual weighing and the smart diaper system in the 108 urination records (A) and the 98 records (B). Mean (mean difference; solid line), 95% limits of agreement (dashed lines), and line of equality (dotted line) are shown.



IAD, Bed Sores, and Adverse Events

A total of 28 participants were evaluated for IAD and analyzed for bed sores (2/30 enrolled participants refused photography). Prior to the start of the study, 8 participants already had bed sores: 6 with grade 2 (partial loss of skin layers) and 2 with grade 3 (necrosis of subcutaneous tissue). At the end of the study, the sores had not increased in size or degree. An improvement from bed sore grade 3 to 2 was noted in one participant. In the remaining 20 participants, de novo IAD or bed sores were not observed. Moreover, during the study period, no adverse events were reported.

XSL•FO
RenderX

User Experience

After the study, caregivers answered a series of survey questions. A total of 22 caregivers completed the questionnaire. The respondents consisted of 12 (54%) employed caregivers, 9 (41%) offspring, and 1 (4%) sibling. The respondents were asked about their willingness to purchase the smart diaper system once it becomes commercialized. About 59% (13/22) and 41% (9/22) had positive and negative responses, respectively. They were also asked about inconveniences related to using the smart diaper system. Cumbersome connections between the diaper and the sensing device or Bluetooth pairing between the sensing device and the smart phone were reported by 6 (27%) of the respondents, and intermittent breakdown of the smart diaper system was reported by 9 (41%; Multimedia Appendix 1).

Discussion

Principal Findings

The objectives of this study were to evaluate the feasibility and utility of the smart diaper system as an alternative to the conventional diaper in a clinical setting. We found that the accuracy of automated urine volume measurement by the smart diaper system compared to the conventional wet diaper weighing method was clinically acceptable. We also found that the smart diaper system had the potential advantage of preventing IAD development and the worsening of bed sores. However, the detection rate of the smart diaper system was low. A large variation in detection rate was observed among participants. Furthermore, we found the inconvenience of the operation and user interface of the smart diaper system presented a problem.

Smart Diaper System in Real Life

This study demonstrated that a smart diaper system developed with IoT technology is applicable in real life, including hospital settings. In acute care hospitals, measuring urine output at regular intervals is often necessary to control the patient's body fluid and electrolyte homeostasis. Therefore, for patients with diapers, health care providers should determine urine volume by counting the number of diaper changes or weighing wet diapers at each change. However, inference based on the number of diapers used is inaccurate and there is often poor compliance for measuring wet diapers. Urinary catheterization is another alternative for more accurate measurement of urine volume, although it is associated with increased urinary tract infections and morbidity. In this study, the average of the absolute values of the difference between urine volume measured by weighing wet diapers on a digital scale and urine volume measured by the smart diaper system was 42.6 mL, which was 14.3% of the average urine volume per urination of 297.2 mL in the 108 paired urination records. Excluding 10 urination records of >500 mL, which was the upper limit of detection of the smart diaper system, the level of difference between the two measurements was only 13.9% (37.2 mL/267.4 mL in the 98 urination records), suggesting clinical acceptability. Consequently, our study suggested that the urine volume measurements obtained using the smart diaper system were more reliable than estimation by counting the number of urinations, more convenient than weighing wet diapers, and less invasive than an indwelling urinary catheter.

Consideration for Low Detection Rate

In this study, the overall voiding detection rate of the smart diaper system was 32.8% (126/384), which was lower than expected. As previously mentioned, considering the small margin of error in urine volume measurement between the two techniques, it was a questionable finding. The detection rate varied widely from participant to participant (range 0%-100%). Figure 3 shows that a large number of voiding detections by the smart diaper system were from a subset of participants, but in 13 participants, less than three counts of voiding detection by the smart diaper system were recorded during the study period. Consequently, we can infer two reasons for this finding. First, when water was poured on the smart diaper pad during the development process, there was a delay of 1-20 minutes in detecting wetness due to the preset time interval of sensing and the Bluetooth connection to save battery power. If the participant had immediately informed the caregiver of urination themselves, or if the caregiver had recognized the participant's urination promptly, diapers could have been changed before detection by the smart diaper system. Second, the caregivers may have had poor understanding and compliance regarding using the smart diaper system appropriately. For the smart diaper system to work well, a good connection between the diaper pad and the sensing device is crucial. Bluetooth pairing between the smartphone and the sensing device is also required. Although a researcher visited and trained the caregiver daily, skillful handling of the first smart diaper system might have been difficult for caregivers. The smart diaper system works well with the researchers' assistance; however, it could be challenge for older caregivers to learn to use the system within the short study period of three days.

Consideration for IAD and Bed Sores

According to the literature, IAD incidence has been reported to be 36% in critically ill patients [25], 7.6% in long-term acute care patients [26], and 5.5% in nursing home residents with new-onset incontinence [27]. IAD may increase the risk of new bed sores, be more frequent in patients with bed sores, and even worsen existing bed sores [28]. In our study, no newly developed dermatitis or bed sores were detected in participants. In addition, aggravation of bed sores was not observed in participants who previously had bed sores, and one case of improvement of bed sores was observed. Timely changing of wet diapers and keeping the area near the perineum dry using the smart diaper system contributed to this effect.

Aged Society and Digital Medicine

As the population ages, the health care expenditures for older adults with multimorbidity are expected to increase steeply. Moreover, increasing socioeconomic gaps can lead to disparities in the health and wellbeing of older people. Digital medicine is an emerging solution for effective control of chronic medical conditions in geriatric populations at a relatively low cost [29]. For the last two decades, eHealth has revolutionized the medical sector with computers, the internet, and electronical medical record systems [30,31]. More recently, the advent of IoT technology and the widespread use of mobile devices have led to the development of mobile health (mHealth), a subdivision of eHealth, which has the potential to improve health outcomes

XSL•FO
RenderX

in chronic disease management [32,33]. However, for now, many older adults and their caregivers have low digital literacy when it comes to using the internet, smartphones, and IoT technology [34]. Therefore, more research about geriatric digital medicine and further development of age-friendly user interfaces are urgently needed to ensure older adults can benefit from digital health.

COVID-19 Pandemic and Digital Medicine

Since the SARS-CoV-2 outbreak started in December 2019, there have been unprecedented impacts across the world. In the era of COVID-19, social and economic activities are changing to non–face-to-face platforms, and health care systems are facing significant challenges related to converting to and using telemedicine. Virtual connections between tertiary hospitals and long-term care facilities were activated in many countries, and "hospital at home," which provides hospital-level care in a patient's home as a substitute for acute hospital care, is emerging as another option in the post–COVID-19 era [35]. In these "contact-free" health care environments, the demand for digital medicine including mHealth is also increasing, and the smart diaper system is expected to play an important role. The smart diaper system could monitor voiding and urine volume, analyze daily or weekly patterns, and transfer the data to medical professionals to help patients at home or in long-term care facilities.

Limitations and Strength

This study had several limitations. First, the study population was small. Second, the study period was only a few days, which might not have provided the caregivers with sufficient time to adapt to the digital equipment. In addition, the study duration may not have been sufficient to observe skin changes in the diaper area, such as IAD and bed sores. Nonetheless, it was a pilot study, and the applicability of the smart diaper system was established. Further large-scale and long-term investigations with age-friendly equipment and user interface are warranted for commercialization and popularization of the smart diaper system. Third, some participants were less compliant, and they did not report any voiding record. A survey conducted after the study showed that 27% (6/22) of the respondents reported inconvenience in using the smart diaper system and 41% (9/22) reported difficulties in operating the smart diaper system. Therefore, developing a more user-friendly system is necessary to improve compliance. Fourth, we did not perform a qualitative and quantitative usability assessment using validated questionnaires, although we reviewed the user experience through a survey after the study.

Nonetheless, the strength of this study is that it provides data on experiences using the smart diaper system in a clinical setting. To the best of our knowledge, this study is the first to investigate the feasibility of a smart diaper system in an acute care hospital. The only previous study was conducted with 18 people with dementia living in nursing homes and estimated the saturation of diaper capacity indirectly rather than measuring the volume of urine [36]. The study reported that the saturation errors between the smart diaper system and FVCs were –26% to 39% in the regular diaper (450 mL) with 51 urination records, and –34% to 30% in the super diaper (1000 mL) with 46 records, respectively. The study did not provide information about the detection rate of the system and concluded that it was not sensitive enough to use as an indicator of the need for a diaper change.

Conclusions

This study suggests that the smart diaper system can promptly notify caregivers of patients' urination to facilitate diaper changes and prevent the occurrence of IAD and bed sores or improve existing IAD and bed sores by reducing exposure time to wet diapers. In addition, the smart diaper system can measure urine volume with reliable accuracy and less invasiveness than the conventional weighing method, reducing the need for routine diaper checks. Therefore, it can save the time and labor of caregivers, as well as reduce patient discomfort. The smart diaper system is expected to help many health care facilities and people in need of care.

Acknowledgments

This study was supported by Panoptics Corp and funded by the National Dementia Research and Development Project from the Ministry of Health and Welfare of South Korea through the Korea Health Industry Development Institute (grant number 1465030924).

Authors' Contributions

S-wK conceptualized and designed the study outline, acquired the data, and supervised the analysis of data. JHC drafted and revised the manuscript. JYC, NHK, YL, JHO, ESK, JR, and KIK supervised the recruitment and data collection, performed analysis, and interpreted the data. JK and YK offered technical support and interpreted the data. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results from a questionnaire on the user experience of the smart diaper system.
[DOCX File , 17 KB - jmir_v23i7e29979_app1.docx]

XSL•FO
RenderX

References

1. Ueda T, Tamaki M, Kageyama S, Yoshimura N, Yoshida O. Urinary incontinence among community-dwelling people aged 40 years or older in Japan: prevalence, risk factors, knowledge and self-perception. Int J Urol 2000 Mar;7(3):95-103 [FREE Full text] [doi: 10.1046/j.1442-2042.2000.00147.x] [Medline: 10750888]
2. Lukacz ES, Santiago-Lastra Y, Albo ME, Brubaker L. Urinary Incontinence in Women: A Review. JAMA 2017 Oct 24;318(16):1592-1604. [doi: 10.1001/jama.2017.12137] [Medline: 29067433]
3. Kim S, Park J, Ahn H, Lee S, Yoo HJ, Yoo J, et al. Risk Factors of Geriatric Syndromes in Korean Population. Ann Geriatr Med Res 2017 Sep 30;21(3):123-130. [doi: 10.4235/agmr.2017.21.3.123]
4. Milsom I, Gyhagen M. The prevalence of urinary incontinence. Climacteric 2019 Jun;22(3):217-222. [doi: 10.1080/13697137.2018.1543263] [Medline: 30572737]
5. Durrant J, Snape J. Urinary incontinence in nursing homes for older people. Age Ageing 2003 Jan;32(1):12-18. [doi: 10.1093/ageing/32.1.12] [Medline: 12540342]
6. Junkin J, Selekof JL. Prevalence of incontinence and associated skin injury in the acute care inpatient. J Wound Ostomy Continence Nurs 2007;34(3):260-269. [doi: 10.1097/01.WON.0000270820.91694.1f] [Medline: 17505245]
7. Karakaya İ, Yenişehir S, Karakaya MG. Nursing Home Residents' Level of Knowledge About Urinary Incontinence. Ann Geriatr Med Res 2019 Mar;23(1):20-26 [FREE Full text] [doi: 10.4235/agmr.19.0003] [Medline: 32743281]
8. Adelmann PK. Prevalence and detection of urinary incontinence among older Medicaid recipients. J Health Care Poor Underserved 2004 Feb;15(1):99-112. [doi: 10.1353/hpu.2004.0001] [Medline: 15359977]
9. Skotnes LH, Omli R, Romild U, Hellzèn O, Kuhry E. Urinary incontinence in Norwegian nursing home residents. OJN 2012 Jun;02(02):116-122. [doi: 10.4236/ojn.2012.22018]
10. Jain P, Parada JP, David A, Smith LG. Overuse of the indwelling urinary tract catheter in hospitalized medical patients. Arch Intern Med 1995 Jul 10;155(13):1425-1429. [Medline: 7794092]
11. Johnson TM, Ouslander JG, Uman GC, Schnelle JF. Urinary incontinence treatment preferences in long-term care. J Am Geriatr Soc 2001 Jun;49(6):710-718. [doi: 10.1046/j.1532-5415.2001.49146.x] [Medline: 11454108]
12. Cruise PA, Schnelle JF, Alessi CA, Simmons SF, Ouslander JG. The nighttime environment and incontinence care practices in nursing homes. J Am Geriatr Soc 1998 Feb;46(2):181-186. [doi: 10.1111/j.1532-5415.1998.tb02536.x] [Medline: 9475446]
13. Gray M. Optimal management of incontinence-associated dermatitis in the elderly. Am J Clin Dermatol 2010;11(3):201-210. [doi: 10.2165/11311010-000000000-00000] [Medline: 20131923]
14. Sugimura T, Tananari Y, Ozaki Y, Maeno Y, Tanaka S, Ito S, et al. Association between the frequency of disposable diaper changing and urinary tract infection in infants. Clin Pediatr (Phila) 2009 Jan;48(1):18-20. [doi: 10.1177/0009922808320696] [Medline: 18648080]
15. Brennan ML, Evans A. Why catheterize?: audit findings on the use of urinary catheters. Br J Nurs 2001 May;10(9):580-590. [doi: 10.12968/bjon.2001.10.9.9389] [Medline: 12066031]
16. Saint S. Clinical and economic consequences of nosocomial catheter-related bacteriuria. Am J Infect Control 2000 Feb;28(1):68-75. [doi: 10.1016/s0196-6553(00)90015-4] [Medline: 10679141]
17. Holroyd-Leduc JM, Sen S, Bertenthal D, Sands LP, Palmer RM, Kresevic DM, et al. The relationship of indwelling urinary catheters to death, length of hospital stay, functional decline, and nursing home admission in hospitalized older medical patients. J Am Geriatr Soc 2007 Feb;55(2):227-233. [doi: 10.1111/j.1532-5415.2007.01064.x] [Medline: 17302659]
18. Kelly JT, Campbell KL, Gong E, Scuffham P. The Internet of Things: Impact and Implications for Health Care Delivery. J Med Internet Res 2020 Nov 10;22(11):e20135 [FREE Full text] [doi: 10.2196/20135] [Medline: 33170132]
19. Fischer M, Renzler M, Ussmueller T. Development of a Smart Bed Insert for Detection of Incontinence and Occupation in Elder Care. IEEE Access 2019 Jul 25;7:118498-118508. [doi: 10.1109/access.2019.2931041]
20. Ziai MA, Batchelor JC. Smart radio-frequency identification tag for diaper moisture detection. Healthc Technol Lett 2015 Feb;2(1):18-21 [FREE Full text] [doi: 10.1049/htl.2014.0098] [Medline: 26609399]
21. Jeong G, Park S. Review of urinary continence care products using sensor technology to improve effectiveness. Proc Inst Mech Eng H 2019 Jan;233(1):91-99. [doi: 10.1177/0954411918794989] [Medline: 30131022]
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Mar 08;1(8476):307-310. [Medline: 2868172]
23. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
24. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995 Sep 15;14(17):1933-1940. [doi: 10.1002/sim.4780141709] [Medline: 8532986]
25. Bliss DZ, Savik K, Thorson MAL, Ehman SJ, Lebak K, Beilman G. Incontinence-associated dermatitis in critically ill adults: time to development, severity, and risk factors. J Wound Ostomy Continence Nurs 2011;38(4):433-445. [doi: 10.1097/WON.0b013e318220b703] [Medline: 21747261]
26. Long MA, Reed LA, Dunning K, Ying J. Incontinence-associated dermatitis in a long-term acute care facility. J Wound Ostomy Continence Nurs 2012;39(3):318-327. [doi: 10.1097/WON.0b013e3182486fd7] [Medline: 22552105]

XSL•FO
RenderX

27. Bliss DZ, Mathiason MA, Gurvich O, Savik K, Eberly LE, Fisher J, et al. Incidence and Predictors of Incontinence-Associated Skin Damage in Nursing Home Residents With New-Onset Incontinence. J Wound Ostomy Continence Nurs 2017;44(2):165-171 [FREE Full text] [doi: 10.1097/WON.0000000000000313] [Medline: 28267124]
28. Foureur N, Vanzo B, Meaume S, Senet P. Prospective aetiological study of diaper dermatitis in the elderly. Br J Dermatol 2006 Nov;155(5):941-946. [doi: 10.1111/j.1365-2133.2006.07423.x] [Medline: 17034522]
29. Kim K. Optimal Management of Chronic Medical Conditions Through Digital Medicine. Ann Geriatr Med Res 2018 Sep;22(3):117-120 [FREE Full text] [doi: 10.4235/agmr.2018.22.3.117] [Medline: 32743260]
30. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
31. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
32. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
33. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
34. Lee H, Park YR, Kim H, Kang NY, Oh G, Jang I, et al. Discrepancies in Demand of Internet of Things Services Among Older People and People With Disabilities, Their Caregivers, and Health Care Providers: Face-to-Face Survey Study. J Med Internet Res 2020 Apr 15;22(4):e16614 [FREE Full text] [doi: 10.2196/16614] [Medline: 32293575]
35. Lekamwasam R, Lekamwasam S. Effects of COVID-19 Pandemic on Health and Wellbeing of Older People: A Comprehensive Review. Ann Geriatr Med Res 2020 Sep;24(3):166-172 [FREE Full text] [doi: 10.4235/agmr.20.0027] [Medline: 32752587]
36. Huion A, Decalf V, Kumps C, De Witte N, Everaert K. Smart diapers for nursing home residents with dementia: a pilot study. Acta Clin Belg 2019 Aug;74(4):258-262. [doi: 10.1080/17843286.2018.1511279] [Medline: 30146971]

Abbreviations

ADL: activity of daily living
FVC: frequency volume chart
IAD: incontinence-associated dermatitis
IoT: Internet of Things
mHealth: mobile health
UI: urinary incontinence

Edited by G Eysenbach; submitted 27.04.21; peer-reviewed by HW Jung, H Mehdizadeh; comments to author 16.05.21; revised version received 29.05.21; accepted 31.05.21; published 30.07.21.

Please cite as:
Cho JH, Choi JY, Kim NH, Lim Y, Ohn JH, Kim ES, Ryu J, Kim J, Kim Y, Kim SW, Kim KI
A Smart Diaper System Using Bluetooth and Smartphones to Automatically Detect Urination and Volume of Voiding: Prospective Observational Pilot Study in an Acute Care Hospital
J Med Internet Res 2021;23(7):e29979
URL: https://www.jmir.org/2021/7/e29979
doi:10.2196/29979
PMID:34328427

©Jae Ho Cho, Jung-Yeon Choi, Nak-Hyun Kim, Yejee Lim, Jung Hun Ohn, Eun Sun Kim, Jiwon Ryu, Jangsun Kim, Yiseob Kim, Sun-wook Kim, Kwang-Il Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Leading Topics in Twitter Discourse on JUUL and Puff Bar Products: Content Analysis

Jon-Patrick Allem[1], MA, PhD; Allison Dormanesh[1], MS; Anuja Majmundar[2], PhD; Vanessa Rivera[1], BA; Maya Chu[1]; Jennifer B Unger[1], PhD; Tess Boley Cruz[1], PhD

[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[2]Department of Surveillance and Health Equity Science, American Cancer Society, Washington, DC, United States

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Department of Preventive Medicine
Keck School of Medicine
University of Southern California
2001 N Soto Street, 3rd Floor, SSB K318
Los Angeles, CA, 90032
United States
Phone: 1 323 442 7921
Email: allem@usc.edu

Abstract

Background: In response to the recent government restrictions, flavored JUUL products, which are rechargeable closed-system electronic cigarettes (e-cigarettes), are no longer available for sale. However, disposable closed-system products such as the flavored Puff Bar e-cigarette continues to be available. If e-cigarette consumers simply switch between products during the current government restrictions limited to 1 type of product over another, then such restrictions would be less effective. A step forward in this line of research is to understand how the public discusses these products by examining discourse referencing both Puff Bar and JUUL in the same conversation. Twitter data provide ample opportunity to capture such early trends that could be used to help public health researchers stay abreast of the rapidly changing e-cigarette marketplace.

Objective: The goal of this study was to examine public discourse referencing both Puff Bar and JUUL products in the same conversation on Twitter.

Methods: We collected data from Twitter's streaming application programming interface between July 16, 2019, and August 29, 2020, which included both "Puff Bar" and "JUUL" (n=2632). We then used an inductive approach to become familiar with the data and generate a codebook to identify common themes. Saturation was determined to be reached with 10 themes.

Results: Posts often mentioned flavors, dual use, design features, youth use, health risks, switching 1 product for the other, price, confusion over the differences between products, longevity of the products, and nicotine concentration.

Conclusions: On examining the public's conversations about Puff Bar and JUUL products on Twitter, having described themes in posts, this study aimed to help the tobacco control community stay informed about 2 popular e-cigarette products with different device features, which can be potentially substituted for one another. Future health communication campaigns may consider targeting the health consequences of using multiple e-cigarette products or dual use to reduce exposure to high levels of nicotine among younger populations.

(J Med Internet Res 2021;23(7):e26510) doi:10.2196/26510

KEYWORDS

electronic cigarettes; JUUL; public health; Puff Bar; social media; Twitter; infodemiology

Introduction

Electronic cigarettes (e-cigarettes) are popular in the United States [1,2]. In February 2020, the US Food and Drug Administration (FDA) restricted flavored closed-system cartridge e-cigarettes (eg, JUUL), with the exception of tobacco and menthol flavors, in an effort to discourage their use among the youth [3,4]. These restrictions did not apply to relatively new disposable (nonrefillable) e-cigarettes [2]. For instance, Puff Bar offers disposable nicotine salt–based products (also

XSL•FO
RenderX

found in JUUL products), in over 20 flavors such as pink lemonade. Congressional lawmakers petitioned the FDA to restrict Puff Bar, arguing that they were the fastest growing e-cigarette brand, replacing JUUL as the e-cigarette of choice among the youth [5]. The FDA sent warning letters to Puff Bar instructing them to remove their flavored disposable e-cigarettes from the market because the latter did not have the required premarket authorization [6]. In response, Puff Bar briefly stopped selling their products through their official website, but their products were always available for purchase from third-party websites.

Recent evidence from Google search trends suggests that public interest in Puff Bars surged immediately after the FDA announced a restriction on flavored e-cigarettes [7]. If consumers simply switch to disposable products during the present restriction on flavored closed-system products, then such restrictions would be less effective. However, it is unclear whether the public views disposable and reusable e-cigarettes as ideal substitutes. In other words, public discourse discussing the product features and user experiences with JUUL and Puff Bar is understudied. A step forward in this line of research is to describe public discourse referencing both Puff Bar and JUUL products in the same conversation.

This study utilized Twitter data to examine public conversations about Puff Bar and JUUL products during a time of change in the e-cigarette marketplace. Twitter has previously been used to describe the context of e-cigarette–related attitudes and behaviors in a way that offers direct insights on user experience, including preferred design features and flavor preferences [8]. By examining the public's conversations about Puff Bar and JUUL products on Twitter, having described themes in posts, this study aims to help the tobacco control community stay informed about 2 popular e-cigarette products with different device features, which can be potentially substituted for one another. Our findings may inform FDA policy targets and communication strategies in the future.

Methods

Posts containing both terms "Puff Bar" and "JUUL" were collected from July 16, 2019, to August 29, 2020, from Twitter's streaming application programming interface (n=2632). Similar to prior studies [8], retweets were removed so that each observation could be treated as an independent observation (n=1577). Two trained researchers manually coded tweets into themes, using an inductive approach. The goal of this approach was to condense the raw text-based data into a summary format and report the underlying patterns that were evident in the data. The unit of analysis was the text. Saturation was determined to be reached with 10 themes.

The codebook (Table 1) consisted of the following themes: (1) device features, including mentions of hardware, product features, specifications, and product quality; (2) flavors, including mentions of flavors offered by each brand or enjoyed by the consumer; (3) longevity, including mentions of how long a Puff Bar or JUUL product lasts, such as the duration and number of puffs; (4) price, including mentions of monetary amounts or affordability of JUUL and Puff Bar products; (5) youth use, including mentions of youth (aged under 21 years) and mentions of children, youth, or teenagers using a Puff Bar or JUUL product or other e-cigarette products during school time or in school premises; (6) switching, including mentions of substituting 1 product with the other; (7) dual use, including mentions of using both Puff Bar and JUUL products; (8) nicotine concentration, including mentions of nicotine concentration or nicotine salt levels; (9) health risks, including mentions of Puff Bar products being more harmful than other e-cigarettes (eg, JUUL e-cigarettes) or vice versa, and of negative health consequences of vaping; and (10) confusion, including mentions of confusion over the differences between the features of Puff Bar and JUUL products. Posts were segregated into multiple themes.

To establish interrater reliability, coders analyzed a subsample of posts (n=300), with agreement ranging from 84% to 97%. The lead author served as the arbitrator and resolved disagreements. Descriptive statistics were reported in a confusion matrix to show the prevalence of each theme as well as theme cooccurrence in a single post. Data collection processes relied on publicly available data and adhered to Twitter's terms and conditions, terms of use, and privacy policies. The protocol was approved by the university's institutional review board (protocol# HS-18-00697).

XSL•FO
RenderX

Table 1. Definitions for each theme and example paraphrased posts.

Theme	Definition	Paraphrased post
Device features	Mentions of hardware (eg, disposable or reusable battery), product features (eg, color of the device), specifications, and quality of the product.	Puff Bar is a disposable, prefilled, and precharged vape, but JUUL has disposable pods and a reusable battery.
Flavors	Mentions of flavors (eg, specific fruit, sweet, savory, candy, alcohol, coffee, tobacco, menthol, or mint) offered by each brand or enjoyed by the consumer.	This banana-flavored Puff Bar e-cigarette tastes amazing compared to a tobacco-flavored JUUL e-cigarette.
Longevity	Mentions of how long a Puff Bar or JUUL product lasts, including the duration and number of puffs.	This Puff Bar only lasts 1 day, but I cannot afford a pack of JUUL pods that last 2 weeks.
Price	Mentions of monetary amounts or affordability of JUUL and Puff Bar products.	Puff Bar [products] cost only US $8 and are cheaper than JUUL [products].
Youth use	Mentions of youth (aged under 21 years) and mentions of children, youth, or teenagers using a Puff Bar or JUUL product or other e-cigarette products during school time or in school premises. Posts may also raise concerns over youth use of vaping products in general.	[I] found a JUUL [e-cigarette] in the high school bathroom in the morning and a Puff Bar [e-cigarette] again later.
Switching	Mentions of quitting 1 product for the other.	I quit [using] JUUL [e-cigarettes], but now I just use Puff Bar [e-cigarettes] every day.
Dual use	Mentions of using both Puff Bar and JUUL products.	[I am] hitting my JUUL [e-cigarette] for breakfast and my pink lemonade Puff Bar [e-cigarette] for dinner.
Nicotine concentration	Mentions of nicotine concentration or nicotine salt levels.	I can get just as much nicotine from Puff Bar [e-cigarettes] as my JUUL [e-cigarette] with even higher nicotine delivery.
Health risks	Mentions of Puff Bar being more harmful than other e-cigarettes (eg, JUUL) or vice versa, and of negative health consequences of Puff Bar products. This may include mentions of people harming themselves by using JUUL or Puff Bar products.	Puff Bar and JUUL [e-cigarettes] made my chest hurt so bad, but I still use my vape.
Confusion	Mentions of confusion over the differences between Puff Bar products and other e-cigarettes (eg, JUUL).	She was holding a Puff Bar or maybe it was a disposable JUUL [e-cigarette]?

Results

The most prominent topic was “flavors” (n=311 of 1577 posts, 19.72%), followed by “dual use” (n=254, 16.11%), “device features” (n=230, 14.58%), and “youth use” (n=219, 13.89%) (Table 2). These were followed by “health risks” (n=130, 8.24%), “switching” (n=105, 6.66%), “price” (n=77, 4.88%), “confusion” (n=49, 3.11%), “longevity” (n=47, 2.98%), and “nicotine concentration” (n=42, 2.66%). The most common cooccurring themes in a single post were “youth use” and “device features” (n=70, 4.44%), followed by “device features” and “flavors” (n=67, 4.25%) and “youth” and “flavors” (n=61, 3.87%).

XSL•FO
RenderX

Table 2. Prevalence of themes[a].

Themes	Flavors	Dual use	Device features	Youth use	Health risks	Switching	Price	Confusion	Longevity	Nicotine concentration
Nicotine concentration	10 (0.63)	1 (0.06)	17 (1.08)	16 (1.01)	4 (0.25)	2 (0.13)	4 (0.25)	0 (0.00)	1 (0.06)	42 (2.66)
Longevity	11 (0.70)	10 (0.63)	7 (0.44)	3 (0.19)	2 (0.13)	4 (0.25)	14 (0.89)	0 (0.00)	47 (2.98)	
Confusion	10 (0.63)	0 (0.00)	3 (0.19)	2 (0.13)	2 (0.13)	0 (0.00)	0 (0.00)	49 (3.11)		
Price	21 (1.33)	4 (0.25)	23 (1.46)	5 (0.32)	2 (0.13)	4 (0.25)	77 (4.88)			
Switching	28 (1.78)	4 (0.25)	5 (0.32)	14 (0.89)	2 (0.13)	105 (6.66)				
Health risks	13 (0.82)	6 (0.38)	9 (0.57)	17 (1.08)	130 (8.24)					
Youth use	61 (3.87)	24 (1.52)	70 (4.44)	219 (13.89)						
Device features	67 (4.25)	17 (1.08)	230 (14.58)							
Dual use	31 (1.97)	254 (16.11)								
Flavors	311 (19.72)									

[a]The diagonal line indicates the prevalence of the 10 topics identified. The off-diagonal lines indicate topic overlap. All values are presented as numbers and percentages in parentheses.

Discussion

Principal Findings

This study provides a summary of public Twitter posts collected over the course of a 13-month period, which includes mentions of both "Puff Bar," a disposable e-cigarette, and "JUUL," a reusable closed-system cartridge e-cigarette. Posts often mentioned flavors, dual use, device features, youth use, health risks, switching 1 product for the other, price, confusion over the differences between products, longevity of products, and nicotine concentration. Theme cooccurrence in a single post was also examined.

"Flavors" was the most common theme in this study, while "flavors" and "device features" represented the second-most common theme cooccurrence in a single post. Prior studies that examined tobacco-related (eg, hookah or little cigars) conversations on Twitter have identified similar themes [9,10]. The FDA has previously taken action to reduce the appeal of e-cigarettes among the youth by removing flavored products. The FDA recently sent warning letters to 10 companies, including Puff Bar, to remove their flavored disposable e-cigarettes from the market because they do not have the required premarket authorization [6]. Puff Bar's compliance with this request and FDA's enforcement will dictate whether their products will be less readily available for purchase.

"Device features" was a predominant theme in this study, while "youth use" and "device features" represented the most common theme cooccurrence in a single post. Previous studies have suggested that product features create lasting psychological, sensory, and behavioral responses among consumers, which may translate to appeal for these products [11]. Additionally, consumers satiate less when similar products are presented as distinct subcategories [12]. In other words, although both JUUL and Puff Bar products are e-cigarettes, consumers may be attracted to using Puff Bar if they perceive it as an e-cigarette product with unique features (such as disposability). Identifying and regulating youth-appealing device features (eg, age restrictions on the purchase of disposable products that are youth-appealing and mandating plain device colors to address attractive designs) may facilitate more effective tobacco control efforts.

Dual use of JUUL and Puff Bar products raises concerns about inadvertent exposure to high levels of nicotine among the youth. A recent study [13] suggests that young adults find it difficult to understand nicotine concentration. When consumers are familiar with both products displaying nicotine levels as mg/mL and percentages, they are more likely to have a correct understanding of nicotine strength [13]. Currently, the official JUUL website and packaging labels list nicotine concentration as percentage values. Similarly, the official Puff Bar website and packaging labels list nicotine concentration as percentage levels; however, this metric appears differently on other retail platforms. As such, regulations standardizing the labeling of nicotine concentration on web-based retail platforms and on product packaging may facilitate consumer awareness. Future health communication campaigns may also consider targeting the health consequences of using multiple e-cigarette products to reduce the dual use of e-cigarette products.

Our findings suggest that there was some level of confusion over the differences or similarities among Puff Bar, JUUL, and other e-cigarettes. Confusion may render Twitter users vulnerable to inaccurate information about the health effects of

XSL•FO
RenderX

these products and likely to misjudge these products' potential relative health risks. A prior study [8] reported that the phrase "What is JUUL?" appeared commonly on Twitter in 2017. Public health communication campaigns need to discuss the health risks of popular emerging products including Puff Bar e-cigarettes as they become increasingly available in the market, to keep parents, educators, and clinicians well-informed of the rapidly evolving e-cigarette marketplace.

Prior studies suggest that marketplaces where consumers can switch to other products in a short period, with limited effort or at a lower price, typically allow easy entry of newer products and facilitate rapid consumer migration to newer products [14,15]. While both JUUL and Puff Bar products contain nicotine salts, Puff Bar potentially facilitates easy switching, given these are single-use products available at a lower cost per unit [16]. Additionally, since Puff Bar is a relatively new e-cigarette brand and its technology could be replicated by other companies easily [17], consumers in the e-cigarette marketplace may transition to other unregulated products. Regulations that create barriers for the entry of similar products and deincentivize consumers to switch to other flavored products are crucial. Currently, flavor restrictions have been applied narrowly to specific product lines, which may make it easier for new products such as those of Puff Bar to circumvent regulations and normalize switching behavior for vape products among consumers.

Limitations

This study was limited to the analysis of discussions on 2 e-cigarette brands, Puff Bar and JUUL, and may not pertain to other e-cigarette brands. However, Puff Barr and JUUL e-cigarettes appear to represent the market leaders for disposable e-cigarettes and reusable closed-system cartridge e-cigarettes, respectively. This study only collected tweets that mentioned the 2 products (Puff Bar and JUUL) in the same post. This decision may have excluded select posts that may have been relevant to our study. This study focused on Twitter posts, and our findings may not generalize to other social media platforms. The posts in this study were collected within a 13-month period and may not extend to other time periods. Data collection relied on Twitter's streaming application programming interface, which prevented the collection of posts from private accounts. Our findings may not be generalizable to all Twitter users or to the population of the United States.

Conclusions

Our findings may offer a point of departure for understanding the public's understanding of and experience with disposable and reusable closed-system cartridge e-cigarettes. Future studies should identify the features of youth-appealing e-cigarette devices to inform more targeted tobacco regulations. Studies should focus on effective communication strategies to raise awareness about known health risks pertaining to dual use and product substitution or switching and about new tobacco products among parents, educators, and vulnerable communities. Comprehensive tobacco regulations may include extending ongoing and upcoming restrictions prospectively to existing and future products, to prevent new products from circumventing current regulations. Regulations mandating standardized labeling of nicotine concentration on web-based platforms may help address health risks from nicotine overdose when consumers switch products. Social media surveillance can help capture new products emerging in the marketplace, such as Puff Bar products, and monitor the web-based marketplace to prevent the sales of nonregulated flavored products.

Acknowledgments

This study was partially supported by funds from the Regents of the University of California, Research Grants Program Office, Tobacco-Related Disease Research Program (grant# T30IR0891). Our study was supported by the National Cancer Institute and the FDA Center for Tobacco Products (grant# U54 CA180905). The funders had no role in study design; collection, analysis, and interpretation of data; drafting of the manuscript; and the decision to submit the manuscript for publication. The opinions, findings, and conclusions herein are those of the authors and do not necessarily represent those of the Regents of the University of California or any of its programs.

Authors' Contributions

JPA had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JPA conceptualized and designed the study. AD, JPA, MC, and VR acquired, analyzed, and interpreted the data. JPA and AD drafted the manuscript. JPA, AD, VR, JBU, AM, MC, and TBC critically revised and provided the final approval for the publication of this manuscript. AD performed the statistical analysis. JPA, JBU, and TBC obtained funding for this study.

Conflicts of Interest

None declared.

References

1. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. e-Cigarette Use Among Youth in the United States, 2019. JAMA 2019 Dec 03;322(21):2095-2103 [FREE Full text] [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
2. Delnevo C, Giovenco DP, Hrywna M. Rapid proliferation of illegal pod-mod disposable e-cigarettes. Tob Control 2020 Dec;29(e1):e150-e151. [doi: 10.1136/tobaccocontrol-2019-055485] [Medline: 32001606]

XSL•FO
RenderX

3. US Food and Drug Administration. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. US Food and Drug Administration. 2020 Jan 02. URL: https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children [accessed 2021-02-02]
4. Kaplan S. Teens Find a Big Loophole in the New Flavored Vaping Ban. The New York Times. 2020 Feb 12. URL: https://www.nytimes.com/2020/01/31/health/vaping-flavors-disposable.html [accessed 2021-02-02]
5. Kaplan S. Lawmakers Say Puff Bar Used Pandemic to Market to Teens. The New York Times. 2020 Jun 02. URL: https://www.nytimes.com/2020/06/02/health/puff-bar-teens.html [accessed 2020-07-13]
6. US Food and Drug Administration. FDA Notifies Companies, Including Puff Bar, to Remove Flavored Disposable E-Cigarettes and Youth-Appealing E-Liquids from Market for Not Having Required Authorization. US Food and Drug Administration. 2020 Jul 20. URL: https://www.fda.gov/news-events/press-announcements/fda-notifies-companies-including-puff-bar-remove-flavored-disposable-e-cigarettes-and-youth [accessed 2020-07-22]
7. Dai H, Hao J. Online popularity of JUUL and Puff Bars in the USA: 2019-2020. Tob Control 2020 Oct 13. [doi: 10.1136/tobaccocontrol-2020-055727] [Medline: 33051277]
8. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Sep 01;190:1-5 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
9. Allem JP, Dharmapuri L, Leventhal A, Unger JB, Boley Cruz T. Hookah-Related Posts to Twitter From 2017 to 2018: Thematic Analysis. J Med Internet Res 2018 Nov 19;20(11):e11669 [FREE Full text] [doi: 10.2196/11669] [Medline: 30455162]
10. Allem JP, Uppu SP, Boley Cruz T, Unger JB. Characterizing Swisher Little Cigar-Related Posts on Twitter in 2018: Text Analysis. J Med Internet Res 2019 Jul 19;21(7):e14398 [FREE Full text] [doi: 10.2196/14398] [Medline: 31325291]
11. Bloch P. Seeking the Ideal Form: Product Design and Consumer Response. J Mark 2018 Nov 28;59(3):16-29. [doi: 10.1177/002224299505900302]
12. Redden J. Reducing Satiation: The Role of Categorization Level. J Consum Res 2008 Feb 01;34(5):624-634. [doi: 10.1086/521898] [Medline: 10.1086/521898]
13. Morean ME, Wackowski OA, Eissenberg T, Delnevo CD, Krishnan-Sarin S. Adolescents and young adults have difficulty understanding nicotine concentration labels on vaping products presented as mg/ml and percent nicotine. Nicotine Tob Res 2021 Jan 12:ntab007. [doi: 10.1093/ntr/ntab007] [Medline: 33433626]
14. Klemperer P. Entry Deterrence in Markets with Consumer Switching Costs. Econ J 1987 Dec 01;97:99. [doi: 10.2307/3038233]
15. Viard V. Do Switching Costs Make Markets More or Less Competitive? The Case of 800-Number Portability. Rand J Econ 2007;38(1):163. [doi: 10.1111/j.1756-2171.2007.tb00049.x]
16. Puff Bar. URL: https://puffbar.com/ [accessed 2020-03-04]
17. Wolfe E. Controversial E-Cigarette Company Puff Bar Says It's Suspending U.S. Sales. Fair Warning. 2020 Jul 13. URL: https://www.fairwarning.org/2020/07/e-cigarette-company-suspends-sales/ [accessed 2020-07-21]

Abbreviations

e-cigarette: electronic cigarette
FDA: US Food and Drug Administration

Edited by R Kukafka; submitted 14.12.20; peer-reviewed by R Sun, OS Liang, MG Kim; comments to author 27.01.21; revised version received 10.02.21; accepted 14.06.21; published 19.07.21.

Please cite as:
Allem JP, Dormanesh A, Majmundar A, Rivera V, Chu M, Unger JB, Cruz TB
Leading Topics in Twitter Discourse on JUUL and Puff Bar Products: Content Analysis
J Med Internet Res 2021;23(7):e26510
URL: https://www.jmir.org/2021/7/e26510
doi:10.2196/26510
PMID:34279236

©Jon-Patrick Allem, Allison Dormanesh, Anuja Majmundar, Vanessa Rivera, Maya Chu, Jennifer B Unger, Tess Boley Cruz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the

XSL•FO
RenderX

Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis

Diana Ramírez-Cifuentes[1], MSc; Ana Freire[1,2], PhD; Ricardo Baeza-Yates[1,3], PhD; Nadia Sanz Lamora[4], MSc; Aida Álvarez[4,5,6], MD, PhD; Alexandre González-Rodríguez[4,5,6], MD, PhD; Meritxell Lozano Rochel[7], BSc; Roger Llobet Vives[7], MSc; Diego Alejandro Velazquez[8], MSc; Josep Maria Gonfaus[9], PhD; Jordi Gonzàlez[8], PhD

[1]Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

[2]UPF Barcelona School of Management, Barcelona, Spain

[3]Institute for Experiential AI, Northeastern University, Boston, MA, United States

[4]Department of Mental Health, Centro de Investigación Biomédica en Red de Salud Mental, Parc Tauli University Hospital, Sabadell, Spain

[5]Parc Tauli Research and Innovation Institute, Sabadell, Spain

[6]Autonomous University of Barcelona, Bellaterra, Spain

[7]Fundación Instituto de Trastornos Alimentarios, Barcelona, Spain

[8]Computer Vision Center, Universitat Autonoma de Barcelona, Bellaterra, Spain

[9]Visual Tagging Services, Bellaterra, Spain

Corresponding Author:
Diana Ramírez-Cifuentes, MSc
Department of Information and Communication Technologies
Universitat Pompeu Fabra
Carrer de Tànger, 122-140
Barcelona, 08018
Spain
Phone: 34 643294184
Email: diana.ramirez@upf.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/10/e33447

Abstract

Background: Eating disorders are psychological conditions characterized by unhealthy eating habits. Anorexia nervosa (AN) is defined as the belief of being overweight despite being dangerously underweight. The psychological signs involve emotional and behavioral issues. There is evidence that signs and symptoms can manifest on social media, wherein both harmful and beneficial content is shared daily.

Objective: This study aims to characterize Spanish-speaking users showing anorexia signs on Twitter through the extraction and inference of behavioral, demographical, relational, and multimodal data. By using the transtheoretical model of health behavior change, we focus on characterizing and comparing users at the different stages of the model for overcoming AN, including treatment and full recovery periods.

Methods: We analyzed the writings, posting patterns, social relationships, and images shared by Twitter users who underwent different stages of anorexia nervosa and compared the differences among users going through each stage of the illness and users in the control group (ie, users without AN). We also analyzed the topics of interest of their followees (ie, users followed by study participants). We used a clustering approach to distinguish users at an early phase of the illness (precontemplation) from those that recognize that their behavior is problematic (contemplation) and generated models for the detection of tweets and images related to AN. We considered two types of control users—focused control users, which are those that use terms related to anorexia, and random control users.

Results: We found significant differences between users at each stage of the recovery process (P<.001) and control groups. Users with AN tweeted more frequently at night, with a median sleep time tweets ratio (STTR) of 0.05, than random control users (STTR=0.04) and focused control users (STTR=0.03). Pictures were relevant for the characterization of users. Focused and

XSL•FO
RenderX

random control users were characterized by the use of text in their profile pictures. We also found a strong polarization between focused control users and users in the first stages of the disorder. There was a strong correlation among the shared interests between users with AN and their followees (ρ=0.96). In addition, the interests of recovered users and users in treatment were more highly correlated to those corresponding to the focused control group (ρ=0.87 for both) than those of AN users (ρ=0.67), suggesting a shift in users' interest during the recovery process.

Conclusions: We mapped the signs of AN to social media context. These results support the findings of previous studies that focused on other languages and involved a deep analysis of the topics of interest of users at each phase of the disorder. The features and patterns identified provide a basis for the development of detection tools and recommender systems.

(J Med Internet Res 2021;23(7):e25925) doi:10.2196/25925

KEYWORDS

social media; Twitter; Spanish; anorexia nervosa; eating disorders; user characterization

Introduction

Background

Mental disorders are psychological syndromes or patterns associated with distress or disability. Anorexia nervosa (AN) is an eating disorder (ED) characterized by the maintenance of a very low body weight, with a constant desire for thinness and a misconception about body shape [1]. EDs, such as AN, are related to risk factors including perfectionistic traits [2], and the existence of comorbid mood disorders such as depression, of which 33%-50% of anorexia patients experience. Anxiety disorders are common in 50% of these patients [3].

The course of AN is highly variable; however, early intervention is strongly associated with full recovery. The phases that patients go through as they overcome AN can be mapped to the transtheoretical model of health behavior change [4], which is described as an integrative method for understanding how people progress toward adopting and maintaining healthy behaviors. This model identifies the following six stages of change: (1) *precontemplation*, where the individual does not know that there is a problem or that a change is required in their life and, thus, does not seek help; (2) *contemplation*, where the person simultaneously considers and rejects the change, while being conscious of the existence of a problem; (3) *preparation*, where the individual starts to take small steps toward behavioral change, believing that it can lead to a healthier life; (4) *action*, in which the person has changed their behavior and intends to sustain it; (5) *maintenance*, a stage wherein the person has maintained the behavioral change for a considerable period (>6 months); and (6) *termination*, a stage wherein the individual has no desire to return to their unhealthy behaviors. It is important to state that relapse, which implies returning from the *action* or *maintenance* stages to an earlier stage, is likely. This approach has been previously evaluated in patients, confirming that the notion of the stages of change is an independent dimension that is very relevant for the treatment of EDs [5].

Automated methods have been designed to detect signs of AN, some of which address the development of early detection approaches [6,7], as it has been proven that the signs and symptoms of mental disorders, including AN, can be traced using social media [6,8-13]. The findings of such research have revealed patterns that can be relevant for the development of tools to detect harmful content [6] and to assist clinicians and psychologists in screening [10-12] and treatment proceedings [6]. Research findings on these topics can also contribute to the improvement of the structure and services provided by online social platforms [6], which are a means through which people with mental disorders can find support for their recovery, as well as they can be used as tools to promote harmful content, which is the case for suicide promoters and pro–eating disorder (pro-ED) communities [6,13].

Recent approaches for characterizing AN on social media account for individual posts from users, meaning that the signs and symptoms are traced by analyzing the content of a single post per user (post level) [6,8,9]. In contrast, there are only a few approaches that address multiple posts of a single user (user level) [10,11]. There are no approaches that address the simultaneous analysis of relational, behavioral, demographical, and multimodal information at the user level, although few studies have combined some of these aspects on AN-related disorders such as anxiety, depression, and suicidal ideation [12,13].

With regard to the type of information extracted and analyzed about EDs on social media, several state-of-the-art approaches are dedicated to the analysis of textual information contained in posts [6,9,14]. They analyze the topics of interest using topic modeling techniques [11], and they also consider the frequency of the terms used through bag-of-words (BoW) models and n-grams [7]. Other textual elements are also considered, such as the frequency of terms closely related to the illness and other lexical and syntactic elements [11,14]. In addition, sentiment analysis tools [15] and methods that use word embeddings, which are vector representations of terms, have been applied [7,16]. Most of these features are later used as representations suitable for the development of predictive models that are often based on machine learning methods [6,7,16]. Through the analysis of text cues, it has been found that users with depression and those with suicidal ideation make more self-references [13], whereas proanorexia and prorecovery communities exhibit distinctive affective, social, cognitive, and linguistic style markers, as proanorexics express greater negative affect, feelings of social isolation, and self-harm [6].

With regard to the analysis of relational information, Wang et al [17,18] analyzed the interactions between users with EDs to identify the topics shared in web-based conversations. In

XSL•FO
RenderX

particular, they explored the interactions among communities with different stances on EDs, considering both pro-ED and prorecovery content, and reported that intercommunity interactions are very limited. In the context of similar disorders, De Choudhury et al [19] constructed egocentric social graphs for postpartum depression detection and found that increased social isolation and reduced availability of social capital on Facebook are the best predictors of postpartum depression detection among mothers. In addition, Colombo et al [20] studied the connectivity and communication of suicidal users on Twitter based on the evaluation of retweets.

Regarding the analysis of images shared by users, only a few studies analyzing the content of images to determine patterns that can be linked to AN signs are available in the literature. This aspect of analysis is important for studies on EDs, as the symptoms of AN, such as extreme weight loss, an obsessive interest in excessive exercise, and restricted food intake through dieting, can be identified from images [1]. Chancellor et al [21] focused on Instagram, which mainly includes image posts, but they exclusively studied textual (captions) and relational elements. Recently, only a few approaches have addressed the actual analysis of image properties to detect mental disorders, such as self-harm, depression, and anxiety [12,22,23]. To the extent of our knowledge, the work of Chancellor et al [16] is the only available study on EDs that analyzed images. Chancellor et al [16] classified moderated pro-ED content and found that the identification of deviant pro-ED content through the analysis of images was possible, reaching a recall of 84%. This approach mainly focused on the detection of harmful images but did not include the analysis of interpretable elements and patterns that might characterize AN-related content on images.

There are also a few studies that consider the stages that patients go through during the course of the illness. De Choudhury [6] and Wang et al [17] addressed the detection and characterization of prorecovery communities, which can be distinguished from pro-ED communities. However, both of these communities did not exclusively include people with the illness. Chancellor et al [24] determined the content and participation measures that could predict the likelihood of recovery of these types of users on Tumblr. They found that only half of the analyzed users were likely to exhibit signs of recovery after a 4-year period, demonstrating the importance of further consideration of the stages that lead to recovery, thus supporting the *transtheoretical model*.

Regarding the analysis of content generated by Spanish speakers with EDs, we identified two approaches that address writings in this language [9,25]. Both approaches are focused on the analysis and detection of users with AN at the post level, with a main focus on textual-based features. Furthermore, these approaches do not address any further recovery stages. Recent works have also addressed other mental conditions such as depression [26] and suicidal ideation [13].

Objectives

This paper addresses the characterization of EDs, in particular AN, in social media. We aim to answer the following research questions (RQs): (1) Which are the visual, textual, behavioral, and demographical elements that characterize and distinguish users at the early, treatment, and full recovery stages of AN on social platforms?; (2) Which elements characterize the social network of users that use anorexia-related terms?; (3) Which are the topics of interest among AN users and their followees?; (4) Which linguistic attributes characterize Spanish-speaking users with AN?; (5) Does the proportion of tweets related to AN significantly change according to the recovery stage?; (6) In a social platform context, can the elements that distinguish users at the precontemplation stage from users at the contemplation stage be identified?; and (7) Are there significant differences between the focused and random control groups?

The main contributions of this work are as follows:

- We generated a Twitter data set that is annotated by psychologists, psychiatrists, and therapists and consists of writings of users manifesting AN (including writings corresponding to treatment and of fully recovered users), and two types of control cases (ie, focused and random control). To the best of our knowledge, this is the first Spanish data set for the analysis of AN at the user level that considers different stages of the illness toward recovery.
- We extracted and inferred several features that consider multiple elements, as follows: images, texts, relations among users, posting patterns, and demographic information. These features were generated to identify elements that characterize users with AN at different stages of the illness and recovery. We also determined the elements that distinguish these users from the two types of control cases.
- We established a deep learning–based method to identify tweets related to AN using individual tweets.
- We trained a predictive model using the images of users with AN and control users to detect whether differences between these groups can be identified on the basis of visual properties.
- We further explored the social network of users with AN through the detection of communities and the analysis of topics of interest of the different types of users, along with those of their followees.
- We applied a clustering approach to distinguish users at the *precontemplation* stage from those at the *contemplation* stage using the *transtheoretical model*.

Methods

Data Collection

We chose Twitter as our main data source, as it has been previously used for studying mental disorders, including EDs, on social media [14,27]. Twitter is suitable as it allows the collection of a set of chronologically organized posts in Spanish and provides metadata relevant for analyzing the relational and behavioral aspects of users.

We selected keywords and popular hashtags commonly used by ED communities, phrases likely to be used by people undergoing treatment, and terms used by recovered users. These keywords and phrases were manually collected and classified from multiple sources in Spanish and English, including proana

XSL•FO
RenderX

blogs, academic publications, and documents made available by the Spanish Association Against Anorexia and Bulimia [14,28,29].

In addition, we conducted a survey among volunteers who have recovered from AN. The phrases and keywords collected were evaluated and filtered by clinicians that were asked to agree on choosing up to 30 keywords or phrases in Spanish that would lead to reach posts from users with AN, including *proana, peso objetivo (objective weight), perder peso (lose weight), IMC (body mass index BMI), sibutramina (sibutramine), mi anorexia (my anorexia), ana y mia (ana and mia)*. We collected 114,627 public tweets containing the search phrases, considering a period in the past, from December 21, 2017, to December 21, 2018. At the same time, a sample of up to 10,000 tweets from the same search period was collected for each user.

To protect the privacy and identity of these users, generic identifiers were assigned to each of them and their posts. In addition, we removed any personal information from the users' descriptions and tweet texts such as usernames, names (proper nouns), email addresses, URLs, password combinations, location names, and all numbers. The extracted metadata elements and texts passed through a strict transformation process to build and exclusively store vector representations of the features of interest, guaranteeing the analysis of fully anonymized information.

Annotation Process

For labeling purposes, we filtered and only considered users with at least three different tweets containing the selected keywords for each category. Among all categories, 645 users met this criterion. Before the submission of the text samples to the annotators, the sample of tweets' texts selected for annotation were anonymized and translated to English to avoid the reidentification of users based on their writings.

We defined five independent groups of users: (1) AN users that manifest the first stages of the disorder and describe signs and symptoms of AN in their texts, which includes users at both the precontemplation and contemplation stages according to the *transtheoretical model*; (2) a focused control group in which, similar to our prior work [13], we included users that did not manifest signs of anorexia but use terms related to the disorder in their writings; (3) treatment users that explicitly stated that they have been diagnosed with AN and are in treatment; (4) recovered users who claim they have recovered from AN; and (5) doubtful cases in which clinicians were not sure about any of the prior categories. A total of 5 annotators participated in the labeling process: 3 psychologists and 2 psychiatrists. These annotators collaborated closely with organizations specializing in the treatment of EDs. The final label for a user's set of tweets was assigned if at least three annotators agreed on the assigned label. For cases where an agreement was not met, the users' tweets were categorized as doubtful cases.

From this first classification approach, a total of 195 users were classified as users with AN, 283 as focused control users, 29 as under treatment users, 18 as recovered users, and 119 as doubtful cases. We performed an interannotator agreement analysis and obtained a Light κ coefficient of 0.4751 (P<.001), which is the result of the averaged Cohen κ values calculated between each pair of annotators. This approach was chosen over Fleiss κ, as all annotators evaluated every sample. The values obtained suggest a moderate agreement among annotators.

In addition to the focused control group, we included another control group consisting of 223 randomly selected users called *random control users*. These users did not necessarily use terms related to AN and were selected using Twitter's Sample Tweets application programming interface (API) [30], which provides a set of random tweets, from which we could obtain the users for our sample. By analyzing the description and a sample of their tweets, we ensured that no users with AN were part of this group. The anonymization and translation process applied for the other groups was followed for this case. Table 1 summarizes the user groups defined initially for the study.

XSL•FO
RenderX

Table 1. User groups defined for the study (N=748).

Group	Description	Collection method	Users, n (%)
AN[a]	• Users most likely at early and advanced stages of AN that do not seem to be on treatment • This group corresponds to the precontemplation and contemplation phases of the transtheoretical model.	AN-related keywords using the Twitter Search API[b]	195 (26.1)
Treatment	• Users most likely in treatment of AN • This group corresponds to the preparation, action, and maintenance phases of the transtheoretical model.	AN- and treatment-related keywords using the Twitter Search API	29 (3.9)
Recovered	• Users that seem to be recovered from AN and have reached the termination stage • They should specify not having a relapse in a long period (≥4 years).	AN- and recovery-related keywords using the Twitter Search API	18 (2.4)
Focused control	• Control users that use vocabulary related to AN, such as psychologists, news accounts, or medical centers.	AN-related keywords using the Twitter Search API	283 (37.8)
Random control	• Twitter users randomly selected with no signs or symptoms of AN	Sample tweets using the Twitter Search API	223 (29.8)

[a]AN: anorexia nervosa.

[b]API: application programming interface.

Characterizing Users With AN on Social Media

This work is dedicated to the characterization of AN considering the multiple stages through which people on the recovery path go through (RQ1), including the analysis of the differences between the two control groups analyzed: focused and random control (RQ7). Our goal is to identify the elements and patterns that distinguish people at each phase based on the analysis of multiple perspectives that are usually considered by clinicians for screening and treatment purposes. The perspectives analyzed include (1) the way Spanish-speaking users express themselves through written posts (content shared) that can lead to the exploration of linguistic dimensions, affective processes and emotions, personal concerns, topics of interest, vocabulary that implies the existence of risk factors associated with the illness, and vocabulary related to signs and symptoms of AN (RQ4); (2) the social network of users, with a focus on the interests shared with their followees and the detection of communities (RQ2 and RQ3); (3) the users' behavior, with the analysis of posting patterns in different periods; (4) demographic aspects such as gender and age ranges; and (5) visual elements that include the analysis of the characteristics of the profile picture of users, along with the images shared on posts.

In addition to the analysis of prior perspectives, for a further exploration of textual cues, we built and evaluated classifiers for the detection of individual tweets related to AN. These tools can be useful for the detection of users with AN, as we assume that a larger number of tweets related to anorexia will be found in the profiles of users living with the disorder. Through this analysis, we also wanted to analyze whether the proportion of tweets related to AN changes significantly according to the recovery stage (RQ5). In addition, these tools can later be useful for automatically filtering writings that are exclusively related to the condition. They can also be used for screening and annotation purposes [13].

We also tested a clustering approach to distinguish cases at the precontemplation stage from cases at the contemplation stage (RQ6). People in the precontemplation phase (also known as the *honeymoon* phase) are characterized by being enthusiastic about their weight loss and the social support they receive. This is a phase in which people believe they are in control of their behavior, and therefore, they are in denial of their unhealthy condition. This behavior differs from the contemplation stage, in which bad habits and symptoms remain, but as the illness progresses, the person notices that there is something wrong and feels that they are losing control of the situation. People at this stage are constantly depressed and may even experience a decrease in their cognitive capabilities [31]. Taking into account these aspects, we consider the detection task relevant, as people at the contemplation stage are more willing to seek help, and therefore, a potential nonintrusive intervention through social media, such as social recommendation, could be successful having users at this stage as a target. However, such potential applications must be further analyzed by considering all the privacy and ethical aspects involved.

Experimental Setup

Data Set Description

The data set built for our analysis consists of a set of features generated and inferred based on the text, images, and metadata of the annotated users' tweets. Table 2 provides relevant information regarding each group in our data collection. For each user, we considered the content from their profiles (tweets) during a year (from December 21, 2017, to December 21, 2018). Some of the initially labeled users were not further considered in the data set, as they had published less than five tweets during the data collection period, which we considered as not informative enough for our analysis purposes. A total of 694 users were part of our final data set, which contained data collected from 2,133,110 tweets, including 405,909 images. It

XSL•FO
RenderX

is important to recall that retweets were not considered within the scope of our analysis.

Table 2. Data set description.

Descriptive item	User group				
	Anorexia nervosa	Treatment	Recovered	Focused control	Random control
Users (n=694), n (%)	171 (24.6)	27 (3.9)	18 (2.6)	271 (39)	207 (29.8)
Tweets collected (n=2,133,110), n (%)	434,615 (20.4)	8317 (0.4)	52,578 (2.5)	1,109,861 (52)	447,739 (21)
Number of tweets collected per user (n=2,133,110), median	1239	1748	2036.5	2608	873
Tweet length (number of words; n=2,133,110), median	14.00	14.00	13.50	12.50	19.00
Images (n=405,909), n (%)	40,142 (9.9)	6584 (1.6)	4202 (1)	298,488 (73.5)	56,493 (13.9)

Comparative Analysis and Evaluation Measures

As part of our data collection process, we extracted, calculated, and inferred some features for performing the analyses required to answer our RQs. For this purpose, we considered network clustering and visualization algorithms (RQ2 and RQ3); prebuilt machine learning models for sentiment analysis; and age range and gender detection tools, including models for the detection of objects in images (RQ1, RQ4, RQ5, RQ6, and RQ7). We also used external sources with lexicons to detect emotions, topics of interest, risk factors, and affective processes lexicons (RQ1, RQ4, RQ6, and RQ7). The description of the extracted features and the calculation or inference processes are detailed in the *Analyzed Perspectives and Features* section. Phyton (Python Software Foundation) was used as our main programming language for data extraction and processing procedures.

To perform a comparison of the groups analyzed, we considered some approaches for hypothesis testing. To do so, we first verified that our numerical features did not follow a normal distribution and that there was no homogeneity of variance for most of them. Therefore, we considered nonparametric tests. First, we applied a Kruskal-Wallis test [32] for multiple independent samples, and after finding features with significant differences among the groups, we proceeded to perform the Mann-Whitney *U* [33] test to check for differences between pairs of groups of interest. As we considered some categorical elements as well, such as age groups, we transformed them into Boolean representations to perform a two-sided proportion *z* test among the groups with these feature types, which is a test equivalent to the proportions chi-square test [34].

Our work addressed three predictive tasks solved using machine learning: one to build a classifier that identifies tweets related to AN (RQ5), another to distinguish users in the precontemplation phase from users in the contemplation phase (RQ6), and the last one to detect individual images related to AN (RQ1). We generated deep learning models based on convolutional neural networks (CNNs) [35] and statistical models such as logistic regression (LR) [36] for the detection of texts and images related to anorexia, taking into account these as supervised tasks for which we have annotated data to learn from. Later, we consider a clustering approach given by a k-means classifier for the detection of users at the precontemplation and contemplation stages, as we did not have data annotated for this purpose. Our predictive approaches were evaluated using machine learning evaluation measures, such as precision, recall, and F1 score. The details of these approaches are described in the *Proportion of AN-Related Tweets, Detection of Precontemplation and Contemplation Phases,* and *Pictures Shared* sections.

Analyzed Perspectives and Features

Overview

In this section, we describe the perspectives and specific features analyzed for the characterization of AN across its stages. The perspectives and categories of the features considered within each perspective are studied by considering their relation to elements that clinicians address for screening and treatment purposes [1,2] and their use for the characterization of related mental health issues such as anxiety [12], depression [23,26], and suicidal ideation [13,20].

Content Shared and Interests

Overview

This perspective addresses RQ1, RQ3, RQ4, RQ5, and RQ7 through an analysis of the textual content shared by users in their tweets. It considers linguistic and psychological aspects through six categories. Some of these categories were based on a classification given by the Linguistic Inquiry and Word Count (LIWC) 2007 Spanish dictionary [37,38], which categorizes words into psychologically meaningful categories. The remaining categories were defined by considering psychological aspects related to EDs, which were defined under the supervision of clinicians. The categories analyzed were as follows: linguistic dimensions (24 features) [37,38], affective processes and emotions (29 features) [37,38], personal concerns and biological processes (12 features) [37,38], vocabulary related to risk factors (10 features) [1,2,13], anorexia-related vocabulary (9 features) [14], and topics of interest to the users (200 topics). From this perspective, we aim to explore the linguistic factors and shared interests that characterize each group. In addition, we map certain elements that are relevant for specialists when identifying signs and symptoms of AN, such as the existence of risk factors, the use of terms related to the illness, and the expression of emotions and personal concerns of the users.

XSL•FO
RenderX

Linguistic Dimensions

These features were extracted using a LIWC dictionary [37,38]. The features belonging to this category address the use of grammatical and syntactical elements such as pronouns, verbs, adverbs, prepositions, and articles, taking into account the different verbal times and pronoun types. These are relevant elements for our analysis because they can give us insight into the elements or people that users are drawing their attention to. These types of elements have been analyzed in prior work where other mental disorders such as depression are analyzed; for instance, Rude et al [39] in their work on depression found that people who are experiencing physical or emotional pain tend to draw their attention to themselves and therefore use more first-person singular pronouns. For our approach, after concatenating all the tweets' texts of a user into one long text, we counted the frequency of words belonging to each of the categories considered from the dictionary, which was normalized by the size (in number of terms) of the resulting concatenation. In addition to the LIWC elements, we also considered the median tweet length (number of terms) as a feature. With the analysis of these features, we also check for the elements that characterize the linguistic attributes of Spanish speakers living with this disorder so that they can be compared with prior studies in English (RQ4).

Affective Processes and Emotions

These features are extracted using the LIWC dictionary and EmoLex [40], which is a dictionary that associates words with eight basic emotions, anger, fear, anticipation, trust, surprise, sadness, joy, and disgust, and two sentiments, negative and positive. This dictionary also provides a file to associate the original terms in English with their Spanish translations. Along with the emotions explored, we addressed cognitive processes, senses, perceptions, and social processes (LIWC). The values obtained for a given user were calculated in the same way as the values for the linguistic dimensions. In fact, the same calculation approach is also used for *personal concerns*, *risk factors*, and *anorexia-related vocabulary* features. The analysis of this aspect is relevant, as general psychiatric disturbance and negative emotionality are elements that characterize people living with EDs [2].

In addition to the prior elements, we make use of a sentiment analysis model, which provides a polarity value for an individual text, in the range of [0,1], from the most negative to the most positive polarity. We used Senti-py [41], trained on Spanish texts from different sources, including Twitter. It is based on a BoW model that goes through an intermediate feature selection process. To obtain a score per user, we calculated the median polarity scores of all the tweets. The higher the polarity score obtained, the more positive the text was expected.

Personal Concerns and Biological Processes

This is another perspective addressed through the analysis of lexicons (LIWC). The elements explored provide a general perspective of the use of terms related to aspects that involve daily activities and concerns, along with biological aspects. Within these aspects, we find terms related to religion; work; leisure; money; and biological processes, such as body, ingestion, and health. These aspects are relevant to our study, as personal background aspects are considered to be relevant for the development of EDs [2]. Biological processes are also relevant to address, as these concerns are representative of patients with anorexia [1]. It is important to specify that we only address the use of general terms such as *God*, *doctor*, or *office*, for instance, we do not do a further exploration regarding personal religious beliefs, professions, or health information within the text.

Vocabulary of Risk Factors

These features correspond to lexicons consisting of up to 3-grams phrases that were mapped to ED risk factors, such as terms referring to self-injuries; suicidal ideation references; self-loathing terms; words that express disdain, substance abuse, lack of social support, family issues; and vocabulary, which refers to discrimination- or abuse-related topics [2,13,42]. The terms and phrases selected for these categories were based on manually mapping common terms and phrases related to anorexia and the terms used in a related task for the detection of suicidal ideation [13].

Anorexia-Related Vocabulary

On the basis of the work of Arseniev et al [14], we used the categories of terms related to AN and its symptoms. Subsequently, we translated them into Spanish. We also kept some of the terms in English, as they are also used by Spanish-speaking users. In addition to these categories, we added names of known laxatives in Spanish [43]. These categories refer to topics such as anorexia promotion, body image, body weight, food and meals, caloric restrictions, compensatory behaviors, and exercise.

Topics of Interest

The topics of interest of a user are also among the elements that we analyze from the content perspective (RQ3), as we would like to know if there is a shift in the main interests of users through the recovery path. For this purpose, we consider predefined topics of interest based on categories that are part of the Empath tool [44], which generates and validates lexical categories on demand from a small set of seed terms (such as football and tennis to generate a sports category). This tool draws connotations between words and phrases by deep learning a neural embedding using a corpus with 1.8 billion words. In our case, 200 prebuilt topics are considered, including sports, music, social media, and politics, among others.

For each user, we calculated the scores for all topics. We take into account that the topics of interest of a user are given by the interests of their followees, the content they like (given by the tweets made by others and marked as favorites), and by the content posted by themselves. For each user, we collected (1) a random sample of their own tweets (up to 500 texts), (2) a random sample of 200 tweets that they had liked during the same period, and (3) the profile descriptions (biographies) of up to 200 random followees of the user. These tweets and descriptions were relevant enough samples of texts that characterized the interests of a user. An individual score with Empath was obtained for each text (tweet or description). Later, the final score for a topic for a given user was calculated by averaging the scores obtained by the topic on all the tweets

XSL•FO
RenderX

considered. It is important to mention that as Empath's categories are in English, we add a translation step before the Empath scores' calculation, using the *Googletrans* Python API [45] for this purpose. The amounts of tweets and descriptions defined for this approach are also based on the request limitations of the API of Twitter and the Googletrans API.

Proportion of AN-Related Tweets

Our RQ5 analyzes whether the proportion of tweets related to AN changes significantly according to the recovery stage, as it is expected that users at the initial stages produce more tweets related to their condition. For this purpose, we first built and compared two models to detect, for each user, if each of their tweets are related to AN. Second, we calculated the median score obtained by the classifier for all user tweets. Finally, we compared the median values for all users belonging to a group to measure the presence of AN tweets in each group. It is expected that users with AN have a median value significantly higher than users in the control and recovered groups. The approach of creating classifiers such as these has previously been used to detect users with suicidal ideations [13]. In this case, the model was built exclusively using a BoW model and LR.

For the AN use case, we trained two classifiers to distinguish tweets of two classes: (1) *anorexia related* and (2) *control*. The instances of the anorexia-related class corresponded to the individual tweets belonging to the users labeled as AN cases (1766 tweets). Later, an equivalent number of tweets was selected to represent the control class; these tweets were randomly extracted using Twitter's Sample Tweets API [30]. The sample of tweets collected was reviewed by annotators to discard those that could possibly belong to the anorexia-related class.

The first classifier was trained over a BoW model with (1-3)-grams at a term level. For this purpose, we used the *Scikit-learn* [46] Python library: *TfIdfVectorizer* to generate a *tf.idf* representation with (1-3)-grams. We considered Spanish stop words [47] and used *ekphrasis* [48] as a text preprocessing tool to replace terms referring to money, hashtags, and emoticons with generic tags. As we obtained 66,404 features, we reduced the number of features to 300 using principal component analysis. We used an LR method and 10-fold cross validation.

For the second classifier, a deep learning approach was applied. The model was defined through a CNN architecture that has been previously applied to text classification tasks [35], including a similar task for suicide risk assessment on social media [13,49]. The same preprocessing approach as that used for the prior model was applied. To train this model, tweets were represented as sequences of terms, and these terms were represented by prelearned word embeddings that were trained over tweets in Spanish [50]. Each tweet was considered as an instance, and its label (anorexia related or control) corresponded to the class assigned to the tweet. For the CNN [35], the embedding sequence instances were given as the model input, where a task-oriented fine-tuning was performed, and we applied a filter window ({2,3,5} terms). We applied max pooling and passed the output to a sigmoid layer to generate the final output. Furthermore, 75% (2649/3532) of the instances were selected for training purposes and the remaining 25% (883/3532) for testing. Among the training instances (tweets), 69.98% (1854/2649) were selected for training the model and 30.01% (795/2649) were considered for validation.

Social Network

Overview

This perspective addresses RQ1, RQ2, and RQ3. Features are extracted taking into account the social network of the user, as elements related to this perspective have been proved to be useful for characterizing other mental conditions such as depression [19] and suicidal ideation [20]. We analyze some features that characterize the user´s popularity and the support received by other users. These features correspond to the number of followers, favorites, and retweets of their posts. We focused on the social network (followees) of users that make use of anorexia-related terms (RQ2), as our goal was to detect communities among these types of users. Furthermore, we explored the likelihood of users with AN to follow users living with the disorder or anorexia promoters by analyzing the topics of interest of their followees. We also explored the differences in their interests and those of the followees of users in treatment and the followees of recovered users.

Measures of Interactions and Engagement

These features are extracted from the metadata of the users' tweets. These features tell us about the relationships and interactions of AN users, which can differ from the interactions of control users [19]. The features extracted and calculated for each user are as follows: number of followees, number of followers, total number of favorites given to the posts of other users, median number of favorites received by the user, and median number of retweets received by the user. These last two features were calculated by considering the tweets of the sample of the user profile.

Analysis of Followees and Community Detection

As part of the perspective that analyzes the social network of a user, we explored the network of users that made use of anorexia-related terms, corresponding to the AN, treatment, recovered, and focused control groups. This was done with the purpose of identifying characteristics of the network that were capable of distinguishing the groups defined, in particular the AN group and the focused control group, as users representing organizations that provide medical and psychological support could be part of it, and it would be relevant to get an insight into the relationships between both groups. For this purpose, we extracted a sample of up to 100 followees of each user from each of these groups (considering Twitter's API request limitations). We built a directed graph where a link between two nodes was given by a *follows* relationship, meaning that users, represented by nodes, are linked to other nodes through directed edges where the arrowheads point to the users they follow. Later, a clustering algorithm was applied to detect communities among these users. We then performed a comparison between the communities automatically detected and what we defined as validation groups, which were created considering the followees of the AN, treatment, recovered, and

XSL•FO
RenderX

focused control groups. These validation groups were defined in such a way that a user was assigned to a validation group (AN, treatment, recovery, or control) if it was mostly followed by users belonging to the originally labeled groups. This was done taking into account up to two followees' levels, denoted as validation subgroups, as explained in Table 3, where we describe the general organization of a group. We considered four main groups and three subgroups per group, where the first subgroup always corresponded to the original users labeled. An instance of a validation group would be *Group AN*, which is composed of three subgroups: G_1 composed of the originally labeled AN's users, G_2 composed of the users mostly followed by G_1 users, and G_3 composed mostly of users followed by G_2.

Table 3. Groups for social network analysis based on users' labels.

Group and subgroup	Nodes user type
Group X[a]	
Subgroup G^b_i	Users manually labeled as part of the X group
Subgroup G_{i+1}	Users mostly followed by G_i
Subgroup G_{i+2}	Users mostly followed by G_{i+1}

[a]X: anorexia nervosa; focused control; treatment or recovered.

[b]G: group.

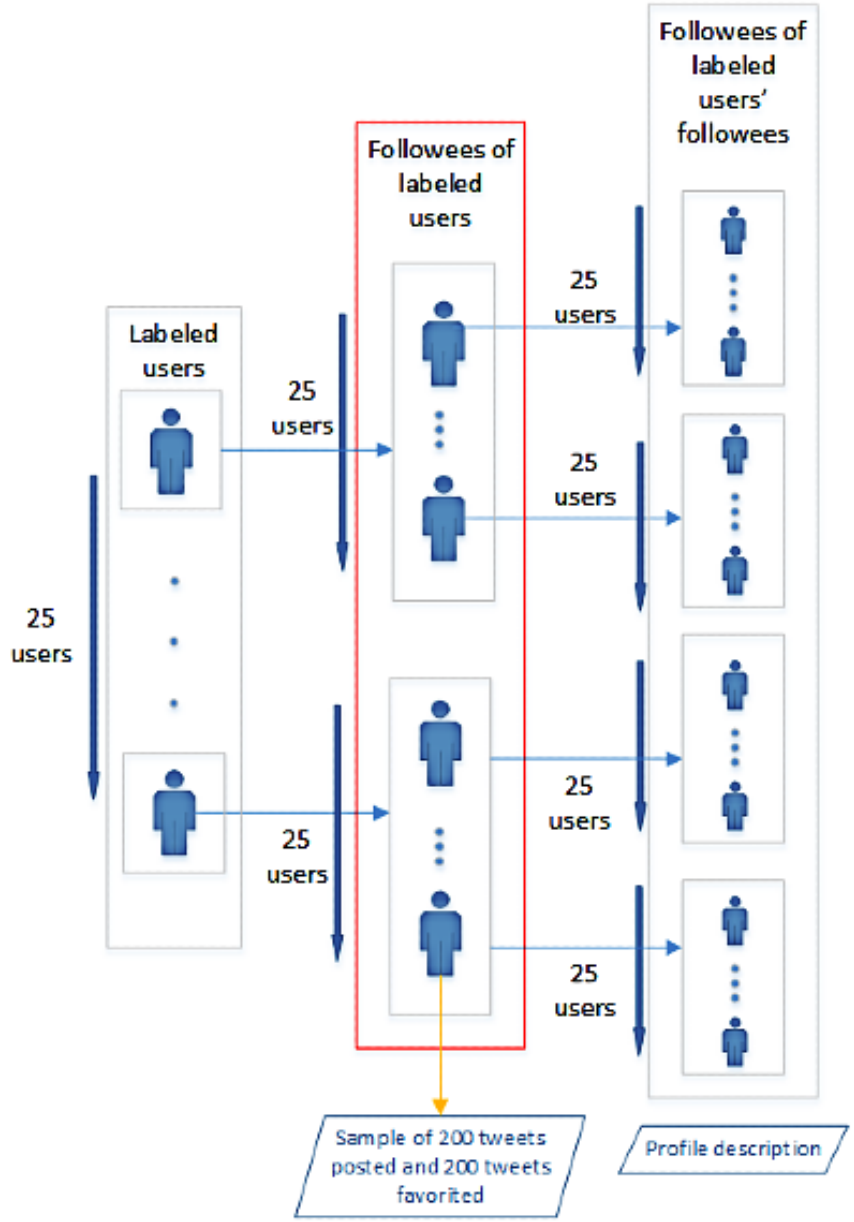
On the basis of a manual revision of a sample (translated to English) of the profile descriptions of users belonging to the communities detected with most nodes, we performed a further analysis of the types of users that were identified as part of each community, and we mapped these communities to our predefined groups so that we could identify which type of users from our groups of interest were part of the communities detected. For the visualization of the social network, we considered the Force Atlas 2 [51] algorithm, and for the detection of communities, we used the method by Louvain [52], both implemented on Gephi [53].

Analysis of Interests Between Users and Their Followees

To address RQ3, as it is our purpose to identify the topics of interest of AN users' followees, we follow the process applied for the analysis of the topics of interest of users of each group, as described in the *Topics of Interest* section, but in this case, we address the followees of each user type. As shown in Figure 1, for this case, we considered up to 25 followees from a sample of up to 25 users per group analyzed. Then, for each of these followees, we calculated scores for the Empath topics by considering the descriptions of 25 random followees, a random sample of their own tweets (up to 200 texts), and a random sample of 200 tweets that they had liked during the same period. The score of a topic for each followee of a user is given again by the average score obtained from all the texts considered. The score for a topic of a user is given by the median of the scores of their followees. Once the scores for the samples of users representing each group were obtained, we calculated the median value of each topic using the scores of each user belonging to the group. Later, we performed a comparison between the interests of users of each group (calculated before) and those of their followees.

XSL•FO
RenderX

Figure 1. Structure defined for the extraction of the interests of the followees of a given user group. For each labeled user of a group, we analyzed the tweets posted and liked by their followees and the profile description of the followees of the labeled users' followees.



Behavioral Aspects

Addressing RQ1 and RQ7 from this perspective, we define features that measure the users' activity to explore elements that can link the frequency of use of social platforms and AN. An instance is the level of activity of users at night, which might imply insomnia, a sign that has been linked to related disorders such as suicidal ideation [13]. These features were extracted from the metadata of the tweets, and they measured the behavior of users based on their activity on a daily, weekly, and monthly basis. The features considered are as follows: the number of tweets on weekdays (Monday-Friday) normalized by the number of tweets from the whole week (weekly tweets ratio), the number of tweets on weekend days (Saturday and Sunday) normalized by the total number of tweets of the week (weekend tweets ratio), the number of tweets posted by the user within each quarter of the year normalized by the total number of tweets of the year period (year quarter tweets ratio), the total number of tweets posted during the year, the median time (in seconds) that passes between the publication of each tweet, and the ratio of tweets posted during the regular sleep time of the user (sleep period tweeting ratio). We also calculated the ratio of tweets posted during the period when the user was usually awake (awake period tweeting ratio). Both of these last features were calculated based on the *sleep time tweets ratio (STTR)* and the *day time tweets ratio (DTTR)* inference approach [13]. This approach was adopted because it addresses the issue of the lack of information regarding the location of the users, which is mostly not shared by them.

For the calculation of the sleep period tweeting ratio, equivalent to the STTR (equation 1) [13], a day is divided into eight fixed time slots of 3 hours each, and it is assumed that an average user has approximately 6 hours of sleep time. This is precisely the period where a fewer number of tweets would be created compared with the rest of the day, so the number of tweets (t) created within each 3-hour time slot is counted for all the tweets of the user, and the sum of the number of tweets within each

XSL•FO
RenderX

pair of continuous time slots is calculated. Then, the minimum score obtained by all the pairs is selected, so that later this value can be normalized according to the total number of tweets of the user (T). It is assumed that the first and last slots are continuous. The awake period tweeting ratio is given by the DTTR, which is equivalent to the difference between 1 and the sleep time ratio: DTTR=1−STTR.



Demographics

We analyzed the demographic characteristics (gender and age features) of the groups to verify whether these correspond to the actual incidence rates [54] (RQ1 and RQ7). These features are inferred, given the fact that Twitter does not publicly display the age and gender of users. We used the approach of Wang et al [55] for demographic inferences. This approach is based on a multimodal deep neural architecture for the joint classification of age, gender, and organizational status of social media users. Their model was trained using data in 32 languages, including Spanish. The method analyzes the description of a user and their profile picture. We used the implementation of the method provided by the authors through a Python library named M3-Inference [55]. The tool outputs scores for three gender categories—male, female, and organization—and four different age ranges.

Before using the detection tool on all the users, to increase its performance, and given the fact that the AN, treatment, and recovered users are not organizations, we defined that only if a user had a score over 0.70, for the organization class, and if this value was higher than the scores for males and females, then this label would be assigned; otherwise, the maximum value among the male and female scores was considered. In addition, if the organization label was assigned to a user, we automatically assigned a specific age group (classified as *an organization*) for all the users classified as organizations. We evaluated the performance of this approach on a group of manually labeled users based on their translated descriptions, where we considered up to 50 users per group. We obtained a macroaverage accuracy of 0.84 for all the gender groups of all the classes and a macroaverage accuracy of 0.80 for all the age groups of all the classes.

Visual Aspects

Overview

In this section, we describe the use of features extracted from the profile pictures of users and from the images of posts shared by users (RQ1 and RQ7). For the first case, we use pretrained models provided by external sources; for the second case, we train a model on our data set and apply it to the pictures shared by a set of validation users from our groups of interest. We explore visual aspects given that there are physical traits that characterize AN [1], and related work has found visual patterns that characterize similar mental conditions [12,13].

Profile Picture

We analyzed 32 features extracted from the pictures of users. We explored the technical features and the detection of emotions and objects. As part of the technical features analyzed, we checked if an image is gray scale, if it is lighter, if it has text, and if it has faces on it. We also analyzed the existence of objects in the pictures. These features are defined through the use of Python libraries such as *imageio* [56] for verifying the brightness of an image, *PIL* [57] for verifying if an image is gray scale, and *the General Recognition AI API* from *Chooch AI* [58], which after taking an image as an input, it outputs the names of elements recognized in the picture, such as texts, clothes, faces, animals, and specific objects. We also detected emotions expressed on the pictures using the Algorithmia facial emotion recognition API, which implements CNN models [59]. The models included in the previous APIs were already trained, so we only ran them over the profile pictures in our data set, and no rights over a further use of these images were granted to the tools' owners. In addition, none of the images were observed by any human annotator, and only the extracted features were stored.

To analyze the existence of objects in the image and represent a user through these features, we considered a Boolean BoW model. In this model, the names of the objects found were the terms considered, and the value for an object was assigned as 1 if it was found on the picture and 0 if it was not found. Given the sparsity of the model, we only considered objects that appeared in at least five images, which led us to retain 20 features of this type. Regarding the emotions' features, we assigned to a user (if there are faces on the picture) the emotion with the highest predictive score, and later, to define a score for a group, we considered the ratio of users assigned to a given emotion. The same approach was considered for the technical features and objects detected.

Pictures Shared

We explored the pictures shared by users through their individual posts to detect AN-related images. For this purpose, we built two models trained on the images shared by users from three of our groups of interest: AN, focused control, and random control. Two binary classification models were trained: (1) an AN versus focused control image detection model and (2) an AN versus random control image detection model. To train the classifiers, the images of a given user were assigned the same label of the user. The method used to build the model was the one explained in the study by Rodriguez et al [60], which is a method for the inference of personality using the OCEAN model, which refers to the big five personality traits: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Moreover, this method has been previously applied to the detection of suicidal ideation–related images [13]. The resulting model was applied to all the images of a set of users kept for validation purposes, which were not considered in the training process. The output of each model was a score for each class to predict. For a single user, the score corresponding to this feature is given by the average score obtained by the classifier for the AN class over all the user's images. For the model generation, a total of 64,615 images were used for training and 15,384 for validation. For the first model (AN vs focused control), 278 users were considered for training and 130 for validation purposes. For the second model (AN vs

XSL•FO
RenderX

random control), 240 users were considered for training and 106 for validation purposes.

The image classifiers consisted of a CNN [61], and we used residual networks [62], which are popular for improving the back-propagation step in deep architectures. For both classifiers, we considered a model that was initially trained on ImageNet [63] and fine-tuned on the target data set (our training images). As in the study by Ramírez-Cifuentes et al [13], we use a 101-layer ResneXt [64] that uses grouped convolution, an architecture that consists of convolution groups of size 32 with a dimensionality of 8 and a fully connected layer at the end, which is the one that performs the classification between our two defined classes: (1) anorexia related and (2) control image (focused or random control depending on the classifier being trained). The training process was performed for eight epochs with stochastic gradient descent with warm restarts [65]. The learning rate was 0.0001, and the weight decay value was 0.001. We used Nesterov with a momentum of 0.99 on two GTX 1080 Ti graphic cards. We also used dropout (50%) to avoid overfitting.

Detection of Precontemplation and Contemplation Phases

According to the labeling approach, within the AN group, we considered users who were probably experiencing AN and did not receive treatment. Among these users, we found those in the precontemplation phase and those at a more advanced stage that might have been willing to start the recovery process (contemplation phase). To detect the elements that distinguish users at the precontemplation stage from those at the contemplation stage, we used a clustering approach. The features selected were those for characterizing users based on their emotions (using the EmoLex features) and that provide an insight on the positivity and negativity of posts as well. We also considered the polarity score, given by the Senti-py library, which classifies a text in the polarity range [0,1]. The use of these features was based on the assumption that people in the precontemplation stage are more positive toward their weight loss and the social support they receive. On the other hand, people in the contemplation phase are at a stage where more negative emotions are manifested due to frustration and lack of control over the weight loss process, which leads to isolation and the affectation of their cognitive capabilities. We used a k-means clustering algorithm for the detection of these communities. Later, we explored the differences among these groups using the values of their features.

Results

Comparative Results of Each Perspective

Overview

In this section, we present the results obtained for all the features analyzed and explore the differences among the values of these features for each defined group. We also explored certain elements within each perspective, reporting on the topics of interest within each group, the generation of the users' social network, and the analysis of features generated using classification models.

From a general perspective, the results suggest the relevance of the analysis of the content (texts) generated by users, for which more attributes with significant differences among the groups were identified. This is consistent with the fact that most of the features map aspects considered by psychologists (eg, emotions, risk factors, and the use of vocabulary that indicates harmful behaviors). Regardless of this, the findings on the social network of users also revealed interesting aspects, such as the evident polarization between the focused control users and AN users. We also found a high similarity between the interests of AN users and their followees. In addition, images are relevant for distinguishing AN cases from control cases. Finally, we identified certain elements that distinguish AN users from treatment and fully recovered users.

Content Shared and Interests

Overview

The results found for this perspective were the most relevant for characterizing AN users (RQ1, RQ3, RQ4, and RQ5). This perspective explored the textual elements from multiple points of view, including linguistic and psychological factors that were particularly useful in distinguishing AN users from control groups. These elements were also important for comparing between our control groups, which were thought to exclusively differ from each other through the use of anorexia-related terms (RQ7). For the majority of the features analyzed for these perspectives, we calculated their median values for each group among all its users. *P* values were also obtained to compare among the following pairs of groups: AN versus treatment, AN versus recovered, AN versus random control, AN versus focused control, and random versus focused control.

Linguistic Dimensions

The results for the 24 linguistic dimension features explored are listed in Table 4. We observed many linguistic features that could distinguish AN users from both the control groups. Notably, the use of first-person singular verbs, and consequently first-person singular pronouns, characterized the posts of AN users, along with a high use of negations and a reduced use of articles. In addition, there were more features with highly significant differences between the AN group and the focused control group (features: 22/24, 92%) than between the AN group and the random control group (features: 15/24, 62%). This can be explained by the fact that, as shown in our further analysis, a high percentage of focused control users were organizations (eg, news sites, nutrition, and medical centers), and their linguistic features were quite distinguishable from those of users with personal accounts. This can also be noticed on the elements that distinguish between random and focused control users, as more personal accounts were part of the random control group.

Regarding the differences between the AN group and users in treatment, we observed significant differences in the use of second-person and first-person plural pronouns, which suggests that there might be a change in their attention focus and a higher level of interaction and inclusion with other people. This pattern was even more evident among recovered users.

XSL•FO
RenderX

Table 4. Comparative analysis among groups based on linguistic dimensions.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
First-person singular verbs	22.2E-03	20.5E-03	17.9E-03	8.14E-03	6.52E-03	.09	<.001[c]	<.001[c]	<.001[c]	.002[d]
First-person plural verbs	9.79E-04	13.8E-04	20.2E-04	17.6E-04	15.5E-04	.01[e]	<.001[c]	<.001[c]	<.001[c]	.03[e]
Second-person singular verbs	3.01E-03	2.90E-03	4.25E-03	2.38E-03	1.99E-03	.34	.008[d]	.001[d]	<.001[c]	.01[e]
Third-person verbs	2.26E-02	2.31E-02	2.50E-02	2.13E-02	1.73E-02	.26	.008[d]	.02[e]	<.001[c]	<.001[c]
Third-person plural verbs	4.99E-03	4.87E-03	5.36E-03	4.45E-03	4.05E-03	.39	.34	.003[d]	<.001[c]	.08
First-person singular pronouns	41.9E-03	41.5E-03	3.0.5E-03	10.1E-03	5.62E-03	.30	<.001[c]	<.001[c]	<.001[c]	<.001[c]
First-person plural pronouns	1.83E-03	2.26E-03	3.75E-03	3.72E-03	3.46E-03	.04[e]	<.001[c]	<.001[c]	<.001[c]	.22
Second-person singular pronouns	9.67E-03	10.6E-03	12.1E-03	8.92E-03	6.52E-03	.16	.003[d]	.24	<.001[c]	<.001[c]
Second-person plural pronouns	2.60E-03	3.38E-03	4.69E-03	2.69E-03	3.12E-03	.004[d]	<.001[c]	.28	<.001[c]	.002[d]
Third-person singular pronouns	3.80E-02	3.93E-02	4.24E-02	4.21E-02	3.91E-02	.05	.005[d]	<.001[c]	.01[e]	.001[d]
Third-person plural pronouns	9.27E-03	9.51E-03	10.6E-03	11.8E-03	10.1E-03	.31	.047[e]	<.001[c]	.06	<.001[c]
Negations	2.66E-02	2.44E-02	2.55E-02	1.98E-02	1.42E-02	.10	.28	<.001[c]	<.001[c]	<.001[c]
Affirmations	7.14E-03	6.46E-03	7.60E-03	7.04E-03	4.50E-03	.33	.24	.19	<.001[c]	<.001[c]
Adverbs	5.24E-02	4.79E-02	4.83E-02	3.81E-02	2.75E-02	.01[e]	.003[d]	<.001[c]	<.001[c]	<.001[c]
Articles	5.67E-02	6.15E-02	6.69E-02	7.06E-02	6.92E-02	.03[e]	<.001[c]	<.001[c]	<.001[c]	.46
Verbs	1.72E-01	1.63E-01	1.71E-01	1.42E-01	1.37E-01	.07	.10	<.001[c]	<.001[c]	.001[d]
Impersonal pronouns	10.9E-02	10.3E-02	10.8E-02	8.28E-02	6.64E-02	.14	.21	<.001[c]	<.001[c]	<.001[c]
Personal pronouns	8.91E-02	9.00E-02	9.00E-02	6.66E-02	5.61E-02	.35	.26	<.001[c]	<.001[c]	<.001[c]
Total pronouns	1.98E-01	1.94E-01	1.94E-01	1.49E-01	1.22E-01	.22	.18	<.001[c]	<.001[c]	<.001[c]
Prepositions	9.81E-02	9.71E-02	10.1E-02	9.97E-02	10.5E-02	.28	.16	.08	<.001[c]	<.001[c]
Past verb tense	1.84E-02	1.88E-02	1.69E-02	1.39E-02	1.18E-02	.43	.11	<.001[c]	<.001[c]	<.001[c]
Present verb tense	1.27E-01	1.22E-01	1.22E-01	1.05E-01	1.00E-01	.13	.17	<.001[c]	<.001[c]	.001[d]
Future verb tense	4.50E-05	7.50E-05	6.70E-05	12.1E-05	12.8E-05	.15	.42	.01[e]	<.001[c]	.07
Median tweet length	14.00	14.00	13.50	12.50	19.00	.45	.43	.001[d]	<.001[c]	<.001[c]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.001.

[d]*P*<.01.

[e]*P*<.05.

XSL•FO
RenderX

Affective Processes and Emotions

The results of these features are described in Table 5. As for the linguistic dimensions, there were significant differences between the values of users of the AN and focused control groups. Negative emotions are found mainly for AN and treatment users; this can be observed also on the expression of emotions such as sadness, disgust, and anger, which are significantly higher for AN users than for control users. Within this same comparison, users with AN use more swearing terms and vocabulary that express anxiety and thoughts on their feelings and perceptions. We observe that there are a few attributes with significant differences between AN and treatment users. For joy and positive emotions (LIWC), the scores were significantly higher for treatment users, which might reflect an improvement in the mood of people as they recover from AN. Regarding recovered users, we also observed the existence of less negative emotions and more positive emotions than AN users. In fact, the expressions of anxiety of recovered users were significantly lower than those of AN users. In addition, their high score on social processes and the highly significant values in comparison with AN users suggest an openness to more interactions with other people. Finally, the differences between random and focused control users are mainly observed through the use of swearing terms, the expression of positive emotions, cause and effects, insight, and discrepancies. In this sense, focused control users seem to be more formal and analytic toward things, which meets the characteristics of accounts that represent organizations. For all the groups analyzed, we can observe in Figure 2 a radar chart that expresses the median values for the eight basic emotions defined by the wheel of emotions by Plutchik [66]. We observed the predominance of sadness over all the other emotions in AN users.

XSL•FO
RenderX

Table 5. Comparative analysis among groups based on effective processes and emotions.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Swearing	15.4E-03	14.6E-03	12.7E-03	8.06E-03	4.33E-03	.48	.19	<.001[c]	<.001[c]	<.001[c]
Absolutist terms	5.36E-03	5.03E-03	5.09E-03	4.98E-03	3.43E-03	.28	.49	.046[d]	<.001[c]	<.001[c]
Positive emotions (Linguistic Inquiry and Word Count)	5.93E-02	6.23E-02	6.50E-02	6.58E-02	5.79E-02	.04[d]	.03[d]	<.001[c]	.14	<.001[c]
Negative emotions (Linguistic Inquiry and Word Count)	6.88E-02	6.50E-02	6.02E-02	4.97E-02	4.44E-02	.24	.01[d]	<.001[c]	<.001[c]	.01[d]
Anxiety	11.2E-03	12.2E-03	8.91E-03	6.61E-03	6.42E-03	.14	.004[e]	<.001[c]	<.001[c]	.36
Cognitive processes	2.73E-01	2.62E-01	2.64E-01	2.23E-01	2.19E-01	.09	.25	<.001[c]	<.001[c]	.25
Cause and effect	1.79E-02	1.72E-02	1.44E-02	1.35E-02	1.62E-02	.13	.03[d]	<.001[c]	.11	<.001[c]
Insight	3.89E-02	3.73E-02	3.72E-02	3.16E-02	3.60E-02	.30	.21	<.001[c]	.003[e]	<.001[c]
Discrepancies	3.99E-02	3.72E-02	3.94E-02	3.03E-02	2.54E-02	.04[d]	.18	<.001[c]	<.001[c]	<.001[c]
Tentative	4.42E-02	4.53E-02	4.33E-02	3.54E-02	3.49E-02	.20	.46	<.001[c]	<.001[c]	.29
Certainty	1.79E-02	1.74E-02	1.99E-02	1.90E-02	1.56E-02	.38	.05	.27	<.001[c]	<.001[c]
Senses and perceptions	5.03E-02	5.01E-02	4.89E-02	3.70E-02	3.80E-02	.48	.24	<.001[c]	<.001[c]	.20
See	1.23E-02	1.19E-02	1.25E-02	1.06E-02	1.11E-02	.30	.16	.02[d]	.06	.19
Listen	10.6E-03	11.5E-03	13.4E-03	8.87E-03	7.19E-03	.04[d]	.002[e]	<.001[c]	<.001[c]	.03[d]
Feel	16.0E-03	17.0E-03	13.7E-03	8.80E-03	9.30E-03	.42	.01[d]	<.001[c]	<.001[c]	.28
Social processes	1.21E-01	1.21E-01	1.40E-01	1.23E-01	1.14E-01	.33	<.001[c]	.05	.18	.01[d]
References to friends	5.50E-03	4.39E-03	8.09E-03	5.06E-03	4.69E-03	.10	.02[d]	.15	.003[e]	.04[d]
References to family	7.97E-03	7.61E-03	9.57E-03	7.93E-03	7.25E-03	.36	.05	.42	.02[d]	.046[d]
Joy	1.44E-02	1.55E-02	1.50E-02	1.35E-02	1.38E-02	.02[d]	.29	.04[d]	.02[d]	.42
Trust	1.90E-02	1.92E-02	1.99E-02	2.27E-02	2.25E-02	.09	.09	<.001[c]	<.001[c]	.42
Fear	1.70E-02	1.76E-02	1.48E-02	1.57E-02	1.59E-02	.25	.07	.008[e]	.02[d]	.25
Surprise	9.36E-03	9.90E-03	9.50E-03	9.07E-03	8.75E-03	.33	.49	.06	.003[e]	.24
Sadness	2.43E-02	2.53E-02	2.24E-02	1.85E-02	1.78E-02	.30	.02[d]	<.001[c]	<.001[c]	.24
Disgust	1.58E-02	1.65E-02	1.41E-02	1.27E-02	1.10E-02	.48	.08	<.001[c]	<.001[c]	.003[e]
Anger	1.51E-02	1.64E-02	1.42E-02	1.39E-02	1.23E-02	.11	.15	.004[e]	<.001[c]	.02[d]
Anticipation	1.48E-02	1.63E-02	1.46E-02	1.51E-02	1.59E-02	.03[d]	.50	.48	.01[d]	.02[d]
Polarity median score	1.86E-01	1.77E-01	1.87E-01	1.94E-01	1.74E-01	.39	.49	.33	.05	.03[d]
Positive emotions EmoLex	3.60E-02	3.73E-02	3.62E-02	3.82E-02	3.97E-02	.07	.32	.046[d]	<.001[c]	.03[d]
Negative emotions EmoLex	3.60E-02	3.72E-02	3.37E-02	3.34E-02	3.02E-02	.42	.07	<.001[c]	<.001[c]	.03[d]

XSL•FO
RenderX

[a]AN: anorexia nervosa.

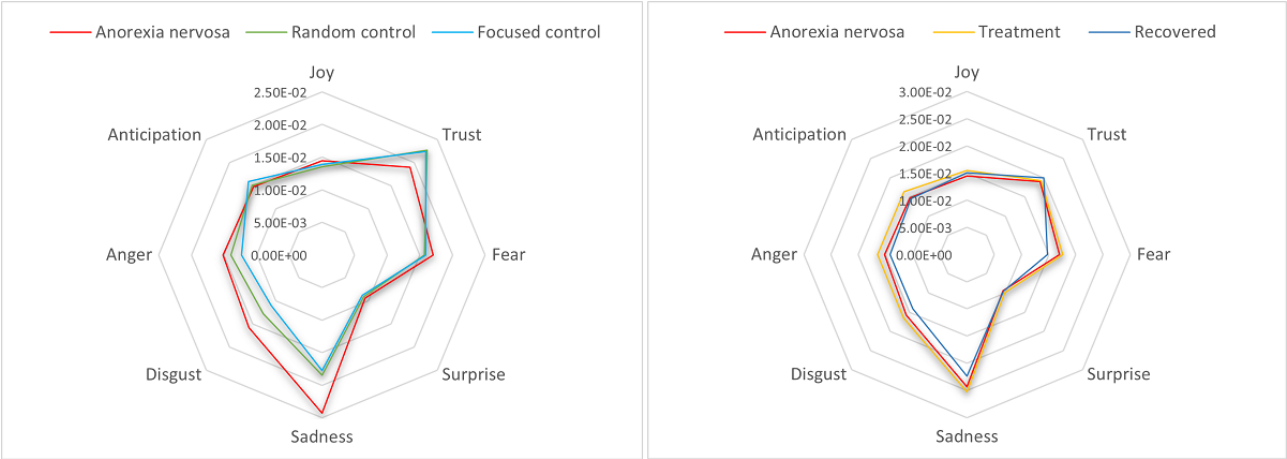
[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.001.

[d]*P*<.05.

[e]*P*<.01.

Figure 2. Comparative scores for emotions according to the wheel of emotions by Plutchik. Left: anorexia nervosa and control groups. Right: anorexia nervosa, treatment, and recovered groups.



Personal Concerns and Biological Processes

The results obtained for these features are listed in Table 6. We observe that most of these features are relevant for distinguishing control users from AN users. Control users discuss more about common concerns such as work, leisure, achievement, money, and religion, whereas AN users are more interested in aspects related to their image, which can be seen by their scores on the body, ingest, health, and biological process categories. There was also a significantly higher interest in death, compared with all other categories. For the treatment group, we observed significantly lower values for the ingest and biological process categories, which might be a sign of improvement in their condition compared with AN users. This is more evident in the comparison of AN and recovered users, where there are very significant differences among the same features. Note that the reference to religious aspects is lower for the AN, treatment, and recovered users in comparison with random control users. Regarding random and focused control users, there are differences in the scores for the body, ingest, health, and biological process categories, as these are the ones that refer to signs of the illness. Focused control users are characterized by their use of AN-related terms, and these findings suggest that among the focused control users, we can find people and organizations that often address the topic of AN. Among these, we can find foundations, medical centers, nutritionists, and psychologists. We later validated this assumption through a social network analysis.

XSL•FO
RenderX

Table 6. Comparative analysis among groups based on personal concerns and biological processes.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Work	3.21E-02	3.50E-02	3.64E-02	4.95E-02	5.25E-02	.05	.01[c]	<.001[d]	<.001[d]	.09
Achievement	3.87E-02	4.08E-02	3.92E-02	4.39E-02	4.30E-02	.07	.48	<.001[d]	<.001[d]	.18
Leisure	1.70e-02	1.91e-02	1.67e-02	1.99e-02	2.12e-02	.16	.40	<.001[d]	<.001[d]	.27
Home	5.52e-03	5.26e-03	6.38e-03	5.20e-03	5.43e-03	.28	.19	.02[c]	.37	.02[c]
Money	8.22E-03	9.42E-03	11.7E-03	13.8E-03	12.2E-03	.03[c]	.006[e]	<.001[d]	<.001[d]	.003[e]
Religion	2.16E-03	2.67E-03	3.42E-03	5.17E-03	3.38E-03	.10	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Sexual	10.6E-03	11.6E-03	15.6E-03	7.98E-03	7.70E-03	.39	.02[c]	.001[e]	<.001[d]	.22
Death	10.5E-03	8.33E-03	6.50E-03	5.87E-03	6.24E-03	.047[c]	<.001[d]	<.001[d]	<.001[d]	.25
Biological processes	9.02E-02	6.86E-02	6.41E-02	3.21E-02	5.01E-02	.04[c]	.003[e]	<.001[d]	<.001[d]	<.001[d]
Body	2.96E-02	2.45E-02	1.92E-02	1.21E-02	1.55E-02	.07	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Ingest	33.8E-03	17.4E-03	15.3E-03	8.52E-03	11.2E-03	.01[c]	.001[e]	<.001[d]	<.001[d]	<.001[d]
Health	17.7E-03	17.5E-03	17.6E-03	6.83E-03	13.2E-03	.50	.47	<.001[d]	<.001[d]	<.001[d]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

Risk Factors Vocabulary

For the use of vocabulary related to risk factors, we noticed that a large number of features were highly significant for the comparison of the AN and control groups. In fact, all the risk factors considered were significant for distinguishing AN from random control users, as shown in Table 7. The use of suicide-related terms is higher for AN users than for all the other groups. Hate and self-loathing terms are found in a lower percentage for recovered users than for AN users and treatment users. We observe that the use of terms related to bullying is higher for treatment and recovered users, which can be explained by the fact that while being on treatment and after recovery, patients are more likely to recognize the issues behind their ED. In general, the scores obtained by all the groups for these features are very low, as these are issues that do not seem to be openly addressed often.

XSL•FO
RenderX

Table 7. Comparative analysis among groups based on vocabulary related to risk factors.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Hate	98.4E-05	68.5E-05	33.1E-05	7.60E-05	2.90E-05	.15	<.001[c]	<.001[c]	<.001[c]	.03[d]
Suicide-related terms	4.20E-05	2.20E-05	0	0	0	.04[e]	.002[e]	<.001[c]	<.001[c]	<.001[c]
Self-harm	1.60E-05	4.60E-05	1.00E-05	0	0	.03[e]	.21	<.001[c]	<.001[c]	.02[e]
Work or school problems	8.80E-05	12.0E-05	6.80E-05	1.60E-05	3.10E-05	.41	.16	<.001[c]	<.001[c]	.12
Self-loathing	4.20E-05	1.90E-05	0	0	0	.25	.003[d]	<.001[c]	<.001[c]	.007[d]
Bullying	0	3.00E-06	11.0E-06	0	0	.06	.03[e]	<.001[c]	.02[e]	<.001[c]
Drugs or alcohol abuse	125E-06	143E-06	77.0E-06	6.00E-06	124E-06	.10	.30	<.001[c]	.40	<.001[c]
Lack of social support	0	2.00E-06	0	0	0	.25	.29	<.001[c]	<.001[c]	.01[e]
Relationship issues	5.80E-05	7.00E-05	7.30E-05	0	1.50E-05	.14	.33	<.001[c]	<.001[c]	.001[d]
Use of antidepressants	0	0	0	0	0	.36	.34	<.001[c]	<.001[c]	<.001[c]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.001.

[d]*P*<.05.

[e]*P*<.01.

Anorexia-Related Vocabulary

These features address the use of vocabulary that describes certain signs and symptoms of AN. The results are presented in Table 8. All the features are highly significant for distinguishing AN users from control cases, and they are all highly significant for distinguishing random from focused control cases. We observed that the scores obtained for the focused control cases were higher than the scores obtained for the random control users. This also happens for the case where recovered and AN users are compared; these users highly differ in the use of vocabulary dedicated to the promotion of AN and vocabulary that expresses concerns regarding body image, body weight, compensatory behavior, and laxatives references, along with caloric restrictions. AN users showed higher scores on these aspects. We also observed that users in the treatment group had lower median values for almost all the features considered, with significant differences in up to four features in comparison with the AN group.

XSL•FO
RenderX

Table 8. Comparative analysis among groups based on anorexia-related vocabulary.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Anorexia promotion	35.4E-04	23.5E-04	13.0E-04	4.99E-04	8.47E-04	.02[c]	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Body image	23.5E-04	7.01E-04	4.26E-04	0	1.60E-04	.01[c]	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Body weight	75.7E-05	32.9E-05	9.00E-05	0	8.90E-05	.11	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Food and meals	29.5E-04	21.5E-04	16.8E-04	1.76E-04	5.20E-04	.16	.02[c]	<.001[d]	<.001[d]	<.001[d]
"Eat" as verb	222E-06	100E-06	91.0E-06	0	9.00E-06	.01[c]	.001[e]	<.001[d]	<.001[d]	<.001[d]
Caloric restriction	443E-06	34.0E-06	2.00E-06	0	0	.001[e]	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Binge eating	3.10E-05	3.40E-05	0	0	0	.40	.004[e]	<.001[d]	<.001[d]	<.001[d]
Compensatory behavior and laxatives	9.00E-04	4.88E-04	2.71E-04	0	0	.09	<.001[d]	<.001[d]	<.001[d]	<.001[d]
Exercise	164E-05	91.8E-05	45.2E-05	7.30E-05	43.7E-05	.05	.001[e]	<.001[d]	<.001[d]	<.001[d]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

Topics of Interest

In this section, we present the results for the exploration of the topics of interest of users that make use of anorexia-related terms (RQ3), which include AN, treatment, recovered, and focused control users. We assume that the interests of random control users are different and depend on the user. This is due to the fact that we do not consider common interest for these users during the data collection process.

Table 9 shows the top 20 topics of interest for the groups according to the Empath categories. We observe that, apart from the elements in common among groups, only users of the AN group refer to topics such as pain, eating, violence, and suffering. Treatment users have many interests in common with AN users, but we can also observe other topics of interest such as reading, music, and sports. Similarly, recovered users rank topics such as sports and weddings in their list. Focused control users also express interest on different topics, with the highest scored topics being health, communication, business, work, internet, and sports, which matches with our prior assumptions regarding this group. Note that for visualization purposes, the actual median values were multiplied by 1000.

To explore the topics of interest in which the groups differed the most from the AN group, we performed the Mann-Whitney *U* test. Figure 3 shows the top 20 topics with the most significantly different values (*P*<.05) between the AN group and the focused control, treatment, and recovered groups. We observe that swearing terms, feminine terms, hate, pain, and appearance obtained high scores for AN users, whereas topics such as economics, college, photography, and work obtained high scores for focused control users. We also observed a limited interest in topics such as music and art for AN users, in comparison with users in treatment, whereas these users (treatment) are less concerned about body and exercise in comparison with users from the AN group. Recovered users are also more concerned about general topics such as law, crime, and politics in comparison with AN users. We report on the percentage of topics found with significant differences among the values for each group (*P*<.05): AN versus focused control, 61% (122/200); AN versus recovered, 40% (80/200); and AN versus treatment, 46.5% (93/200). We also calculated the values for Spearman rank correlation coefficient based on the median values obtained for each topic in each group. The following pairs of groups were compared: AN versus recovered (ρ=0.87), AN versus treatment (ρ=0.87), AN versus focused control (ρ=0.67), treatment versus focused control (ρ=0.87), recovered versus focused control (ρ=0.87), and treatment versus recovered (ρ=0.97). We observe that AN and focused control users are less interested in similar topics, whereas treatment and recovered users' interests are more correlated with those of the focused control group.

XSL•FO
RenderX

Table 9. Top 20 topics of interest (using Empath) among groups that use anorexia-related vocabulary and their median values.

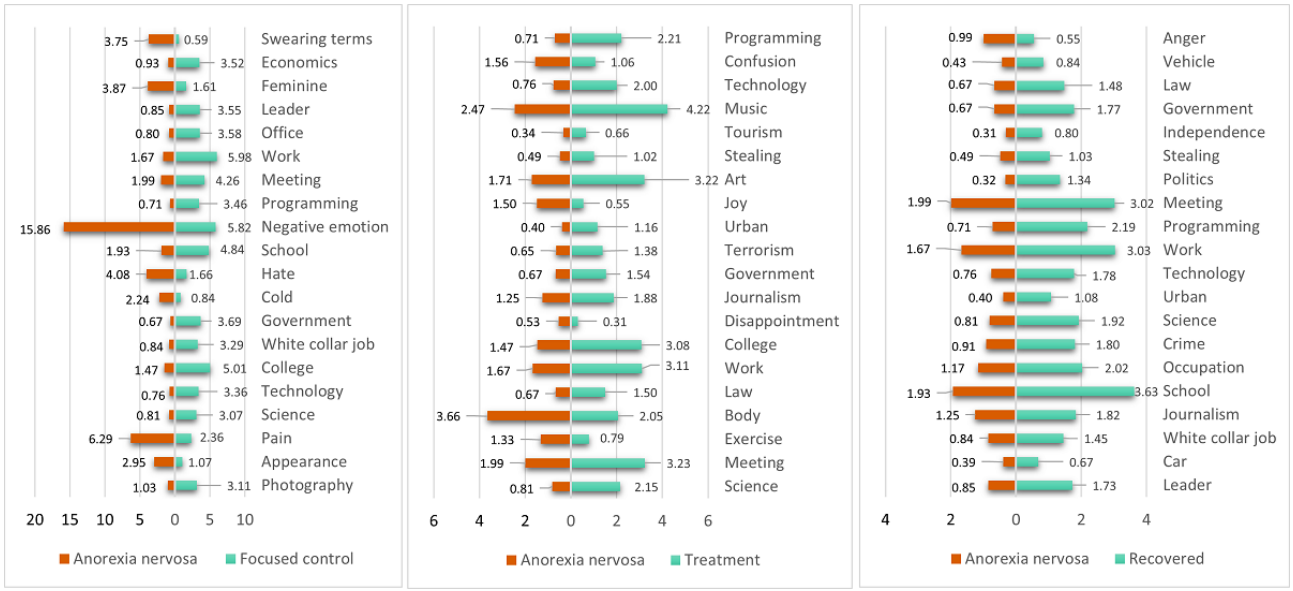
Groups and topics	Value, median
Anorexia nervosa	
Negative emotion	15.86
Friends	8.24
Speaking	7.67
Positive emotion	7.27
Children	6.41
Pain	6.29
Eating	6.19
Communication	6.13
Optimism	5.93
Family	5.91
Love	5.60
Shame	5.40
Violence	5.21
Party	4.84
Social media	4.70
Suffering	4.25
Home	4.24
Hate	4.08
Childish	4.06
Feminine	3.87
Treatment	
Negative emotion	10.50
Friends	7.28
Positive emotion	6.95
Speaking	6.74
Social media	6.62
Children	6.55
Communication	6.06
Optimism	5.61
Family	5.23
Party	4.85
Love	4.80
Reading	4.51
Music	4.22
Home	4.10
Internet	4.09
Musical	3.99
Listen	3.94
Wedding	3.79
Violence	3.63
Sports	3.62

XSL•FO
RenderX

Groups and topics	Value, median
Recovered	
Negative emotion	10.85
Friends	9.78
Speaking	8.09
Positive emotion	7.60
Communication	7.01
Children	6.66
Family	6.46
Social media	5.62
Home	4.95
Party	4.95
Optimism	4.91
Love	4.91
Eating	4.31
Wedding	3.95
Sports	3.94
Giving	3.94
Violence	3.89
Childish	3.75
Pain	3.75
Affection	3.71
Focused control	
Health	9.39
Communication	7.41
Business	6.90
Work	5.98
Positive emotion	5.88
Internet	5.83
Negative emotion	5.82
Social media	5.71
Speaking	5.54
Sports	5.44
Messaging	5.03
College	5.01
Eating	4.92
Children	4.86
School	4.84
Family	4.59
Reading	4.41
Party	4.30
Optimism	4.28
Meeting	4.26

XSL•FO
RenderX

Figure 3. Top 20 topics with most significantly different values (P<.05) between the anorexia nervosa group and the focused control, treatment, and recovered groups. The median values for each feature are shown.



Proportion of AN-Related Tweets

Regarding RQ5, the results obtained by the LR classifier at the training approach (cross validation) were as follows: F1 score=0.97, precision=0.98, and recall=0.97, whereas the results for the deep learning approach averaged after multiple runs over the test set were as follows: precision=0.98, recall=0.98, and F1 score=0.98. As the second model obtained slightly better results, it was applied to all the tweets of all the users regardless of the group they belonged to. For each user, the value considered as a feature was the median score obtained by the classifier on all tweets. We then compared the median values of each group analyzed.

We used the Mann-Whitney U test to perform an analysis of the median score provided by the classifier to all the users' tweets. We applied the classifier to all groups of users. The median values for each group are the following: AN (0.23), treatment (0.13), recovered (0.08), random control (0.03), and focused control (0.05). The P values for the group comparisons are the following: AN versus recovered (P<.001), AN versus treatment (P=.004), AN versus focused control (P<.001), treatment versus focused control (P<.001), and focused control versus random control (P=.02). We noticed very significant differences between the AN group and all other groups considered. Notably, the median classifier score obtained by AN users was higher than that obtained by users from all other classes. Moreover, the median values for the groups decreased according to the recovery stage, meaning that the score was lower for recovered users than for treatment users. Note that focused control users obtain a higher score than random control users, as focused control users address AN-related topics.

Social Network

Overview

Our findings from this perspective (RQ2 and RQ3) provide an insight into the users that are part of the focused control group and verify that organizations and specialists are part of it. We also observed that the structure of the social network could tell us about the different types of accounts that make use of AN-related terms, in particular, through the use of clustering approaches. Our results also showed a high polarization between focused control users and AN users, which was verified by the analysis of the topics of interest between the followees of each user group, where there were significant differences in 75% (150/200) of the analyzed topics.

Measures of Interactions and Engagement

The results obtained for these features (Table 10) show that focused control users have a significantly higher median number of followers and followees than AN users. The median number of followers of these users (focused control) shows that these accounts have a higher number of followers than random control users, which might be an indicator of the popularity of these user types that are more likely to be organizations. We also observe that AN users have a reduced number of interactions with other users in comparison with treatment, recovered, and random control users (based on the favorites given). In general, we observe that a reduced number of tweets generated by all user groups are liked or retweeted by other users, probably because they consume this type of information in a discrete way or because they do not generate very popular content.

XSL•FO
RenderX

Table 10. Comparative analysis among groups based on interaction and engagement measures.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Number of followers	621.50	815.00	600.00	540.00	1174.00	.02[c]	.26	.26	<.001[d]	<.001[d]
Number of followees	286.50	483.50	289.50	492.00	509.00	.02[c]	.18	<.001[d]	<.001[d]	.24
Given favorites	7746.50	10,893.00	23,955.00	10,085.50	4917.00	.04[c]	.004[e]	.02[c]	.02[c]	<.001[d]
Received favorites	0.00	1.00	0.50	0.00	1.00	.11	.22	.001[e]	.001[e]	<.001[d]
Received retweets	0.00	0.00	0.00	0.00	0.00	.04[c]	.29	.07	<.001[d]	<.001[d]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

Analysis of Followees and Community Detection

As explained in the *Methods* section, we analyzed the structure of the social network of users using AN-related vocabulary (RQ2). In Table 11, we report on the percentages of nodes belonging to each group defined through the approach previously explained in Table 3. Most of the users considered were part of the focused control group, followed by AN, recovered, and treatment users. For visualization of these groups, we used Gephi, as shown in Figure 4. A total of 99,283 nodes were considered, with each node representing a user. The average number of edges per node (average degree of the graph) was 2.57, the shortest distance between the two most distant nodes in the network (full network diameter) was 15, and the average path length was 4.72, which represents the average number of steps it takes to get from one member of the network to another. The average clustering coefficient was 0.017, which implies that most of the nodes were not related. To ease the visualization and interpretation of the results, we applied a *k*-core filter with *k*=2 to see the maximal subgraph with a minimum degree equivalent to *k*. The number of nodes displayed in Figure 4 is 12,680, and the size of the nodes is given by the page rank score obtained by each node. The graph clearly shows the polarization between the AN and focused control groups, with the few treatment and recovery cases displayed in between and closer to the focused control cases.

XSL•FO
RenderX

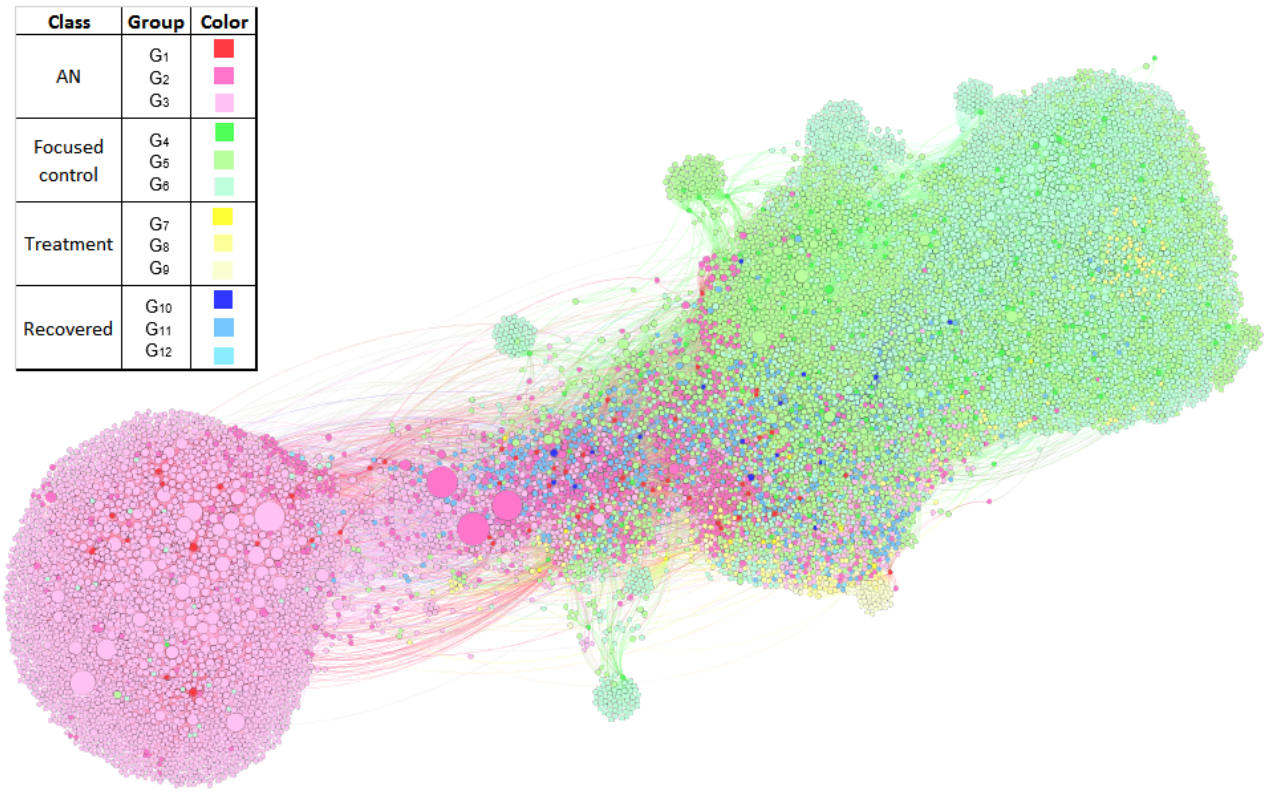
Table 11. Graph information of subgroups according to the type of followers.

Validation group and subgroup	Nodes user type	Node, %
AN[a]		
G^b_1	Labeled as AN	0.05
G_2	Mostly followed by G_1	10.55
G_3	Mostly followed by G_2	12.46
Focused control		
G_4	Labeled as focused control	0.16
G_5	Mostly followed by G_4	46.56
G_6	Mostly followed G_5	20.57
Treatment		
G_7	Labeled as treatment	0.01
G_8	Mostly followed by G_7	3.95
G_9	Mostly followed G_8	0.74
Recovered		
G_{10}	Labeled as recovered	0.02
G_{11}	Mostly followed by G_{10}	4.63
G_{12}	Mostly followed G_{11}	0.3

[a]AN: anorexia nervosa.

[b]G: group.

Figure 4. Visualization of the social network of the AN, focused control, treatment, and recovered groups according to the types of users they are mostly followed by. Each group is represented by a different color. Groups associated with the same class have similar colors. AN: anorexia nervosa; G: group ID.



XSL•FO
RenderX

For further analysis of the full network, we applied a clustering algorithm to detect the communities within it. We found 80 communities and obtained a modularity value of 0.86. As shown in Table 12, we analyzed the descriptions (biographies) of the users of the 10 communities with the highest node percentages. We describe the types of users found in each community and identify the types of users from our annotated groups that are part of each community. Figure 5 shows the network, with the automatically identified communities highlighted. For comparison, we used the same structure displayed in Figure 4. It can be seen that the community with the highest number of nodes is GC_1, which corresponds to the community of users that are likely to have an ED and users that might be anorexia and bulimia promoters (they correspond to big nodes in the graph, ie, higher page rank, meaning more popular nodes). We also observe two other relevant communities that mainly correspond to focused control cases (GC_2 and GC_3) and are characterized by having users that represent organizations and specialists on mental health issues and nutrition centers. We also observe a community that corresponds to news and TV accounts (GC_5), which also characterizes focused control users. We see that the small number of treatment and recovered users are part of different communities that address multiple topics and that the communities that gather users from different groups are those that involve singers, artists, influencers, and leisure-related topics. These results show that users at the precontemplation and contemplation stages are isolated from accounts that offer assistance to overcome the illness. In this sense, recommender systems might enforce this behavior of the network because they tend to recommend a user to follow similar accounts.

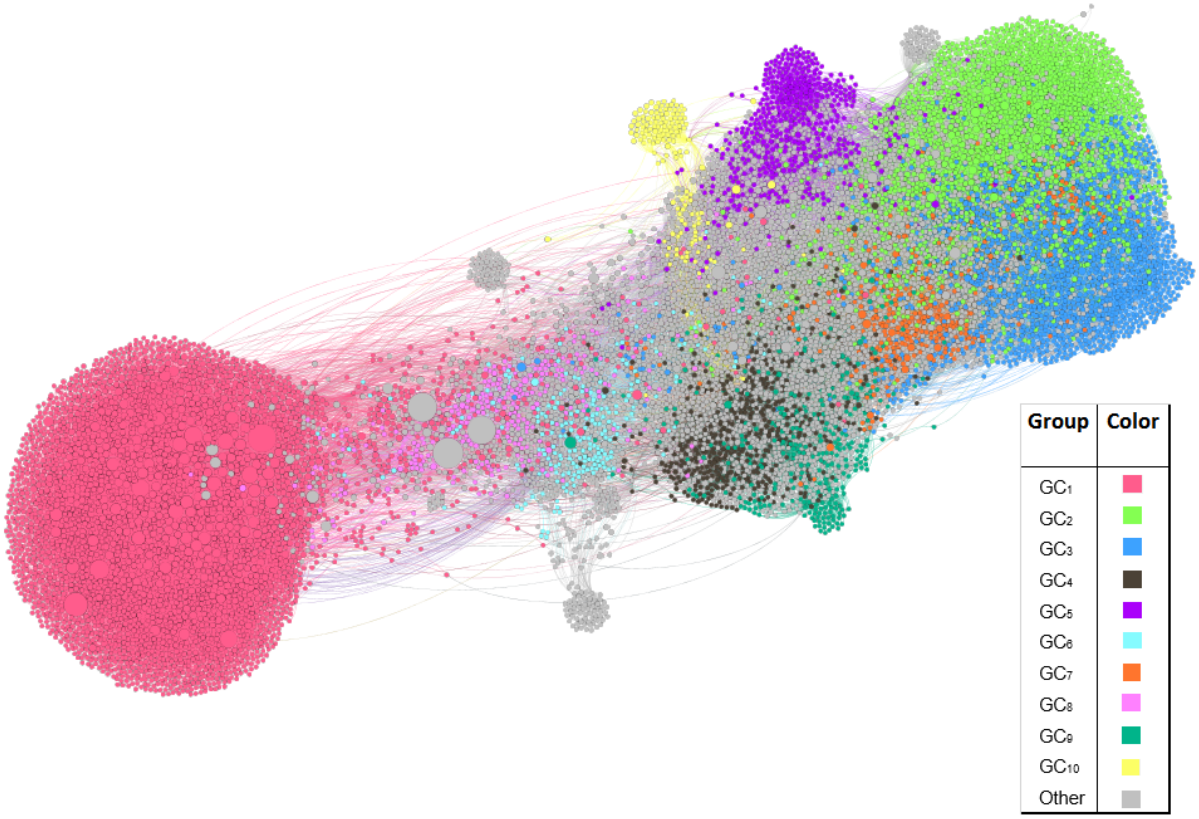
Table 12. Description of the types of users identified in each community with the highest node percentages.

Community	Description of the community	Group users identified	Node, %
GC^a_1	Users with eating disorders and anorexia and bulimia promoters	Anorexia nervosa	9.65
GC_2	Organizations, medical centers, and psychologists	Focused control	8.04
GC_3	Nutritionists and nutrition centers	Focused control	7.37
GC_4	Varieties: influencers	Focused control	3.88
GC_5	News and television	Focused control	3.72
GC_6	Fans of pop singers	AN and recovered	2.69
GC_7	Undefined varieties	Treatment and focused control	2.54
GC_8	Undefined varieties	Recovered and AN	2.53
GC_9	Comics, anime, and drawing	Treatment and focused control	2.31
GC_{10}	Uruguayan community	Focused control	2.29

[a]GC: group community identifier.

XSL•FO
RenderX

Figure 5. Visualization of the 10 communities with the highest node percentages. Each group corresponds to a community and is represented by a given color. GC: group community.



Analysis of Interests Between Users and Their Followees

Regarding the results for the topics of interest of the users who make use of anorexia-related terms (RQ3), Table 13 shows the top 20 topics of interest for our groups' followees according to the Empath categories. We observe that negative emotions, eating, pain, death, and violence are among the topics most relevant to AN followees. Regarding the other groups, we cannot observe a pattern that would normally characterize each user type; instead, we observe interest in all types of topics, which is more evident in focused control users. We can also observe that topics such as friends, family, children, and parties are relevant for most of the groups.

For a better comprehension of the results on this topic analysis task, we explored the topics in which certain followee groups differ the most. We used the Mann-Whitney *U* test for this purpose. Figure 6 shows the top 20 topics, with the most significantly different values ($P<.05$) between the AN followees group and the focused control, treatment, and recovered followee groups. We see a very high value for negative emotion on AN followees in comparison with focused control followees. Appearance is also a topic in which AN followees differ from focused control followees and recovered followees. We also report on the percentage of topics found with significant differences among the median values for the following pairs of groups ($P<.05$): AN followees and recovered followees, 45% (90/200); AN followees and focused control followees, 75% (150/200); AN followees and treatment followees, 48% (96/200); AN and AN followees, 22% (44/200); and recovered and recovered followees, 21% (42/200). From this, we observe that AN users and recovered users differ the least in their interests with their own followees. AN followees and focused control followees show the biggest difference in interests. We also calculated the values for Spearman rank correlation coefficient based on the median values obtained for each topic in each group. The pairs of groups were as follows: AN and AN followees ($\rho=0.96$), treatment and treatment followees ($\rho=0.97$), recovered and recovered followees ($\rho=0.96$), focused control and focused control followees ($\rho=0.97$), AN followees and treatment followees ($\rho=0.93$), AN followees and recovered followees ($\rho=0.93$), AN followees and focused control followees ($\rho=0.69$), treatment followees and focused control followees ($\rho=0.86$), and recovered followees and focused control followees ($\rho=0.86$). From these results, we can say that for all the groups, their interests are highly similar to those of their followees; however, the interests of the treatment and recovered followees groups are more highly correlated to the focused control group followees than the AN followees group, indicating a change in interest through the evolution of the disorder.

XSL•FO
RenderX

Table 13. Top 20 topics of interest and their Empath median values for the groups that use anorexia-related vocabulary followees.

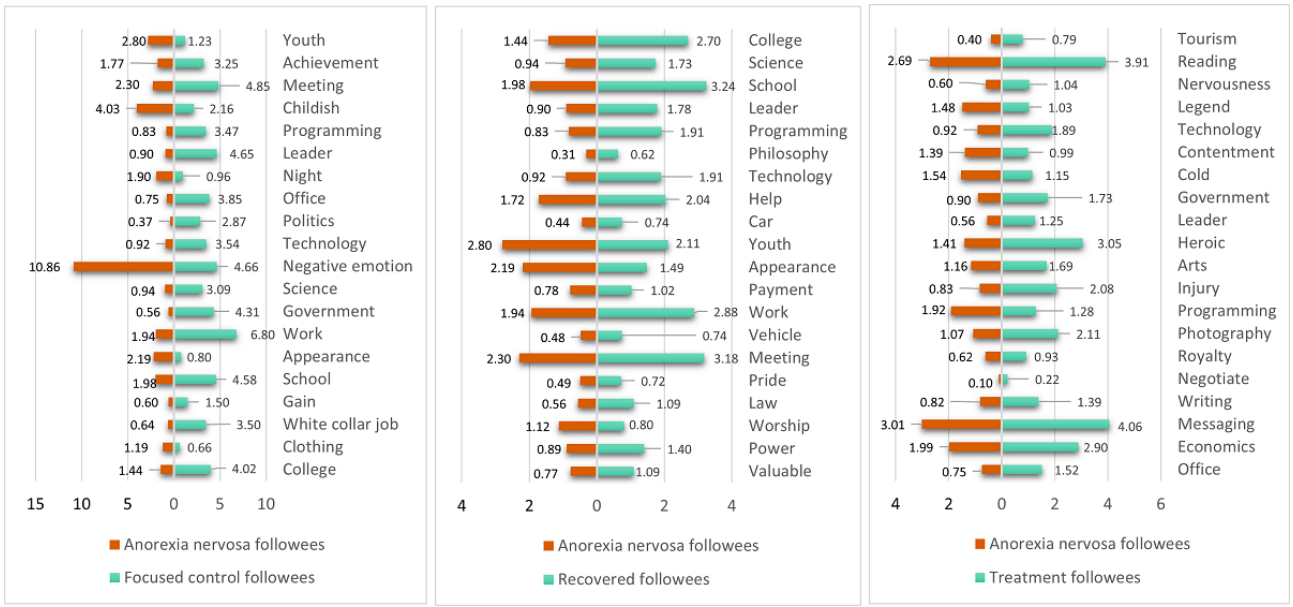
Groups and topics	Value, median
Anorexia nervosa	
Negative emotion	10.86
Friends	7.58
Positive emotion	7.30
Speaking	6.50
Communication	5.93
Optimism	5.85
Children	5.80
Social media	5.71
Party	5.62
Love	5.09
Family	4.69
Childish	4.03
Giving	4.00
Eating	3.93
Home	3.83
Pain	3.72
Death	3.72
Wedding	3.70
Violence	3.62
Celebration	3.42
Treatment	
Negative emotion	7.92
Friends	7.76
Positive emotion	7.02
Communication	6.96
Social media	6.83
Speaking	6.42
Children	5.48
Party	5.43
Optimism	5.22
Family	4.82
Love	4.28
Internet	4.15
Music	4.09
Messaging	4.06
Listen	4.02
Musical	3.92
Reading	3.91
Wedding	3.76
Celebration	3.74
Childish	3.72

XSL•FO
RenderX

Groups and topics	Value, median
Recovered	
Negative emotion	9.73
Friends	7.71
Positive emotion	7.23
Communication	7.15
Speaking	6.87
Social media	5.69
Optimism	5.55
Party	5.22
Children	5.20
Family	4.98
Internet	4.33
Giving	4.00
Love	3.96
Messaging	3.83
Reading	3.73
Wedding	3.70
Celebration	3.67
Listen	3.57
Home	3.52
Childish	3.47
Focused control	
Business	7.29
Communication	7.04
Work	6.80
Positive emotion	6.41
Internet	5.72
Social media	5.36
Party	4.86
Meeting	4.85
Speaking	4.79
Negative emotion	4.66
Leader	4.65
Reading	4.60
School	4.58
Messaging	4.57
Children	4.56
Occupation	4.50
Family	4.49
Optimism	4.35
Government	4.31
Celebration	4.12

XSL•FO
RenderX

Figure 6. The top 20 topics with most significantly different values (P<.05) between the anorexia nervosa followees group and the focused control, recovered, and treatment followees groups. The median values for each feature are shown.



Behavioral Aspects

The results of the behavioral aspects (RQ1) analyzed (Table 14) showed that AN users tweeted more on weekends compared with control groups. In addition, the median time between tweets was lower for AN users (they tweeted more frequently) in comparison with random and focused control users. We also observed that the tweeting ratio during sleeping periods was significantly higher for AN users than for the control groups. This might indicate some sleep alteration, which is a usual sign in EDs and other associated mental issues, such as depression. Regarding the tweeting periods during the year, we see that between December and February (winter in Europe and summer in most countries of South America) users from the AN group tweeted less than users from all the other groups. However, we cannot match this finding to a clinical fact related to the seasons of the year, given the lack of information regarding the users' location. Figure 7 shows the comparative box plot of the values for each group on the sleep period tweeting ratio and the weekend tweet count ratio.

XSL•FO
RenderX

Table 14. Comparative analysis among groups based on behavioral aspects.

Features	AN[a], median	Treatment, median	Recovered, median	Random control, median	Focused control, median	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Working week tweet count ratio	0.73	0.73	0.75	0.75	0.75	.37	.15	.03[c]	<.001[d]	.01[c]
Weekend tweet count ratio	0.27	0.27	0.25	0.25	0.25	.37	.15	.03[c]	<.001[d]	.01[c]
Median time between tweets	625.25	701.00	1187.50	4063.75	1088.00	.43	.15	<.001[d]	.005[e]	<.001[d]
Sleep period tweeting ratio	0.05	0.05	0.04	0.04	0.03	.30	.07	<.001[d]	<.001[d]	.001[e]
Awake period tweeting ratio	0.95	0.95	0.96	0.96	0.97	.30	.07	<.001[d]	<.001[d]	.001[e]
Normalized tweet count per year quarter: December to February	0.01	0.16	0.15	0.18	0.16	<.001[d]	.02[c]	<.001[d]	<.001[d]	.01[c]
Normalized tweet count per year quarter: March to May	0.01	0.23	0.21	0.23	0.21	.001[e]	.11	<.001[d]	.01[c]	.001[e]
Normalized tweet count per year quarter: June to August	0.27	0.32	0.27	0.25	0.24	.39	.29	.10	.004[e]	.06
Normalized tweet count per year quarter: September to November	0.36	0.24	0.27	0.28	0.31	.002[e]	.20	.001[e]	.32	.002[e]
Number of tweets created since the account creation	7910.00	21,038.50	18,409.00	23,291.50	21,463.00	.004[e]	.12	<.001[d]	<.001[d]	.49

[a]AN: anorexia nervosa.

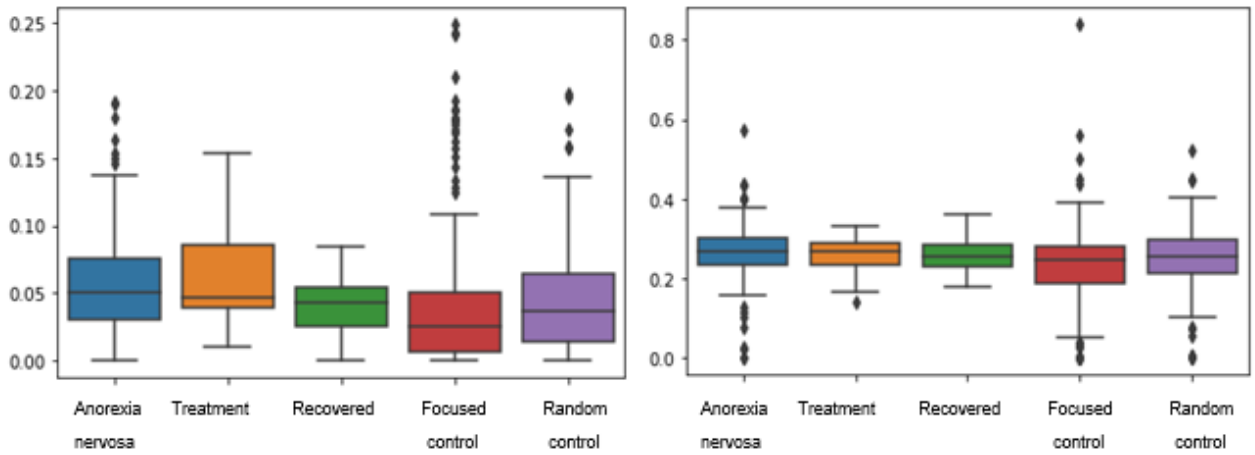
[b]*P* values were analyzed using Mann-Whitney *U* test.

[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

Figure 7. Comparative box plots of the sleep period tweeting ratio (left) and the weekend tweet count ratio (right) of each user group.



XSL•FO
RenderX

Demographics

Using the age and gender inference tool, we obtained the percentages of users corresponding to demographic categories (RQ1) for each group (Figure 8). Most of the AN, treatment, and recovered users were young women. These results are compatible with the statistics that mention that the incidence rates for AN are the highest for women aged 15-19 years [54]. We also observe that a considerable number of users in the focused control group represent organizations. When comparing the ratios of users belonging to each gender per group (Table 15), we see differences between the AN and control groups due to the number of female users. We also observed differences between the focused control and random control groups, as there were fewer organizations in the random control group. Regarding age (Table 16), we observe that AN users differ from the control groups because of the large number of AN users aged ≤18 years. We also find differences between the AN and recovered groups, as recovered users are normally older than AN users. This is consistent with the fact that a full recovery process often takes years, and therefore, users get older as the recovery stages are reached.

Figure 8. Composition of the anorexia nervosa, treatment, recovered, and control user groups according to gender and age. Each age and gender subgroup is represented by a color.

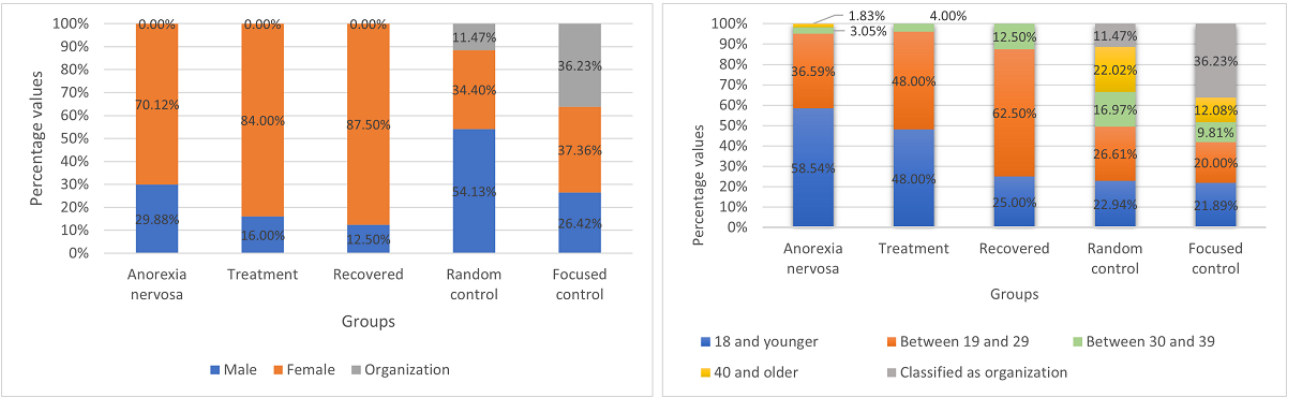


Table 15. Comparative analysis among groups based on gender.

Gender	AN[a] versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Male	.15	.14	<.001[c]	.44	<.001[c]
Female	.15	.14	.04[d]	<.001[c]	.50
Organization	N/A[e]	N/A	<.001[c]	<.001[c]	<.001[c]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using proportions *z* test.

[c]*P*<.001.

[d]*P*<.05.

[e]N/A: not applicable.

Table 16. Comparative analysis among groups based on age, including users classified as organizations.

Age groups, years	AN[a] versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
≤18	.32	.01[c]	.002[d]	<.001[e]	.78
19-29	.27	.04[c]	.31	<.001[e]	.09
30-39	.80	.06	<.001[e]	.009[d]	.02[c]
≥40	.50	.59	<.001[e]	<.001[e]	.003[d]

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using proportions *z* test.

[c]*P*<.05.

[d]*P*<.01.

[e]*P*<.001.

XSL•FO
RenderX

Visual Aspects

The visual aspects analyzed address the profile picture of the user and the images shared on their posts (RQ1). We found significant differences between the groups' profile pictures. Regarding the technical aspects (Table 17), we can observe that focused control users are likely to be distinguished from AN users because of the presence of text in their profile pictures. This also applies to random control users, who tend to use text as well but in a lower ratio than focused control users. These findings can be explained by the use of logos in the profile pictures of the accounts of organizations. In addition, AN users' pictures are significantly darker than focused control users' pictures. In terms of the emotions detected (Table 18), we observed that treatment users expressed more neutral emotions. We also notice that sadness characterizes AN users, which are the only ones with a ratio of users showing such expressions. On the objects detected (Table 19), there were significant differences in the existence of clothing elements between the AN and control groups, along with the appearance of hands, shorts, and accessories, which might suggest that more full-body pictures are shared by AN users, which might consequently imply a higher interest in their appearance. We also observe that there is a high ratio of posters on the control users' profiles, which validates our prior assumption about the representation of organizations. We also observed that few men were identified on pictures of users of the AN group, whereas women were identified on more than half of the AN profile's pictures.

Table 17. Comparative analysis among groups based on technical aspects of profile pictures.

Features	AN[a] ratio of users	Treatment ratio of users	Recovered ratio of users	Random control ratio of users	Focused control ratio of users	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Is gray scale	0.09	0.13	0.17	0.05	0.03	.78	.57	.31	.06	.43
Is lighter	0.50	0.63	0.50	0.42	0.65	.51	.99	.30	.03[c]	<.001[d]
Has text	0.06	0.13	0.00	0.22	0.34	.51	.53	.008[e]	<.001[d]	.045[c]
Has faces	0.28	0.50	0.33	0.25	0.26	.21	.79	.66	.79	.82

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using proportions *z* test.

[c]*P*<.05.

[d]*P*<.001.

[e]*P*<.01.

Table 18. Comparative analysis among groups based on emotions detected from profile pictures.

Features	AN[a] ratio of users	Treatment ratio of users	Recovered ratio of users	Random control ratio of users	Focused control ratio of users	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Neutral	0.00	0.38	0.00	0.05	0.09	<.001[c]	N/A[d]	.06	.02[e]	.34
Sad	0.09	0.00	0.00	0.00	0.00	.37	.43	.002[f]	<.001[c]	N/A
Fear	0.05	0.00	0.17	0.03	0.02	.53	.23	.61	.20	.43
Surprise	0.02	0.00	0.00	0.00	0.01	.72	.76	.22	.61	.39
Angry	0.09	0.13	0.00	0.06	0.07	.78	.43	.46	.56	.83
Happy	0.03	0.00	0.17	0.08	0.09	.61	.12	.18	.12	.80
Disgust	0.00	0.00	0.00	0.03	0.02	N/A	N/A	.15	.32	.43

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using proportions *z* test.

[c]*P*<.001.

[d]N/A: not applicable.

[e]*P*<.05.

[f]*P*<.01.

XSL•FO
RenderX

Table 19. Comparative analysis among groups based on objects detected from profile pictures.

Features	AN[a] ratio of users	Treatment ratio of users	Recovered ratio of users	Random control ratio of users	Focused control ratio of users	AN versus treatment, *P* value[b]	AN versus recovered, *P* value[b]	AN versus random control, *P* value[b]	AN versus focused control, *P* value[b]	Random versus focused control, *P* value[b]
Poster	0.00	0.00	0.00	0.09	0.24	N/A[c]	N/A	.01[d]	<.001[e]	.004[f]
Clothing	0.75	1.00	0.50	0.53	0.33	.11	.19	.005[f]	<.001[e]	.003[f]
Person	0.28	0.13	0.00	0.32	0.12	.34	.13	.58	.007[f]	<.001[e]
Man	0.05	0.00	0.00	0.32	0.16	.53	.59	<.001[e]	.03[d]	.003[f]
Dress	0.03	0.00	0.17	0.01	0.02	.61	.12	.34	.74	.47
Boy	0.02	0.00	0.00	0.02	0.01	.72	.76	.81	.61	.40
Tree	0.02	0.00	0.17	0.05	0.03	.72	.03[d]	.23	.53	.43
Human hand	0.06	0.00	0.00	0.00	0.01	.47	.53	.01[d]	.02[d]	.39
Fashion accessory	0.08	0.00	0.00	0.02	0.02	.41	.48	.08	.03[d]	.77
Flower	0.03	.0.00	0.00	0.01	0.04	.61	.66	.34	.79	.19
Glasses	0.00	0.13	0.00	0.05	0.03	.004[f]	N/A	.06	.16	.43
Animal	0.00	0.00	0.00	0.02	0.02	N/A	N/A	.25	.22	.90
Shorts	0.06	0.13	0.00	0.00	0.00	.51	.53	.01[d]	.004[e]	N/A
Jeans	0.03	0.00	0.00	0.02	0.01	.61	.66	.68	.21	.40
Human eye	0.06	0.00	0.00	0.01	0.01	.47	.53	.06	.02[c]	.83
Cat	0.00	0.13	0.00	0.01	0.02	.004[f]	N/A	.41	.22	.47
Footwear	0.09	0.00	0.00	0.05	0.02	.37	.43	.31	.01[d]	.12
Human nose	0.02	0.00	0.00	0.02	0.00	.72	.76	.81	.16	.10
Girl	0.19	0.25	0.17	0.0	0.03	.67	.90	.006[f]	<.001[e]	.43
Woman	0.58	0.75	0.67	0.18	0.22	.35	.67	<.001[e]	<.001[e]	.38

[a]AN: anorexia nervosa.

[b]*P* values were analyzed using proportions *z* test.

[c]N/A: not applicable.

[d]*P*<.05.

[e]*P*<.001.

[f]*P*<.01.

Finally, when it comes to the exploration of the pictures posted by users, we developed two classifiers: (1) one trained with the images of users from the AN group and users from the focused control group and (2) another trained with the images of users from the AN group and users from the random control group. The results showed that there were highly significant differences between the AN and control groups (*P*<.001 for both comparisons). The median of the aggregated scores of the first classifier for the AN class (AN vs focused control) for a set of 130 validation users was 0.73, whereas the median value for focused control users was 0.36. This means that a higher number of pictures related to AN were found on the posts of AN users. When analyzing the median of the aggregated scores of the second classifier (AN vs random control), on a set of 106 validation users, the median value for AN users was 0.78 and for random control users was 0.54. We observed lower aggregated scores for both control cases, meaning that these users share fewer AN-related pictures. These results show that the pictures are informative for the detection of users with AN.

Regarding the visual content of the images posted by users with AN, there are some visual patterns that emerge after training CNNs, as described previously. These visual patterns were found to be quite characteristic in such posted images because they received the highest scores by the CNN. As expected, most images correspond to body objectification, that is, selfies, extremely thin body parts (mostly legs), and (altered) images of unrealistic ideals of body size. There are also several images of scales showing (rather low) weight numbers. Finally, and less frequently, we can identify rather healthy food images, mostly salads and fruits.

XSL•FO
RenderX

Detection of Precontemplation and Contemplation Phases

After performing the clustering approach with k-means, we obtained two clusters of users among those labeled as AN (RQ6). We analyzed both groups and found strong indicators of the groups identified as our contemplation (cluster 1) and precontemplation (cluster 2) groups. This was mainly because the precontemplation cluster was characterized by higher values of positivity compared with its negativity, whereas the opposite occurred for the contemplation cluster (Figure 9). In addition, the median polarity score for the precontemplation cluster (0.1915) was significantly higher (P<.001) than that for the contemplation (0.1694) cluster. In fact, there were significant differences (P<.001) in the values for all features considered in the clustering approach. In addition, the differences between opposite (positive and negative) emotions such as joy and sadness are lower for the values of the precontemplation cluster, as the early stages of the illness are characterized by signs of the person feeling in control of the weight loss process and being enthusiastic about their progress and the social support received. The median values for all features analyzed for each cluster are displayed in Table 20. Among the AN users, a total of 115 users were assigned to the precontemplation cluster, and 56 users were assigned to the contemplation cluster.

Figure 9. Polarity (calculated using Senti-py) and positivity and negativity (calculated using EmoLex) values for the detected precontemplation and contemplation clusters.

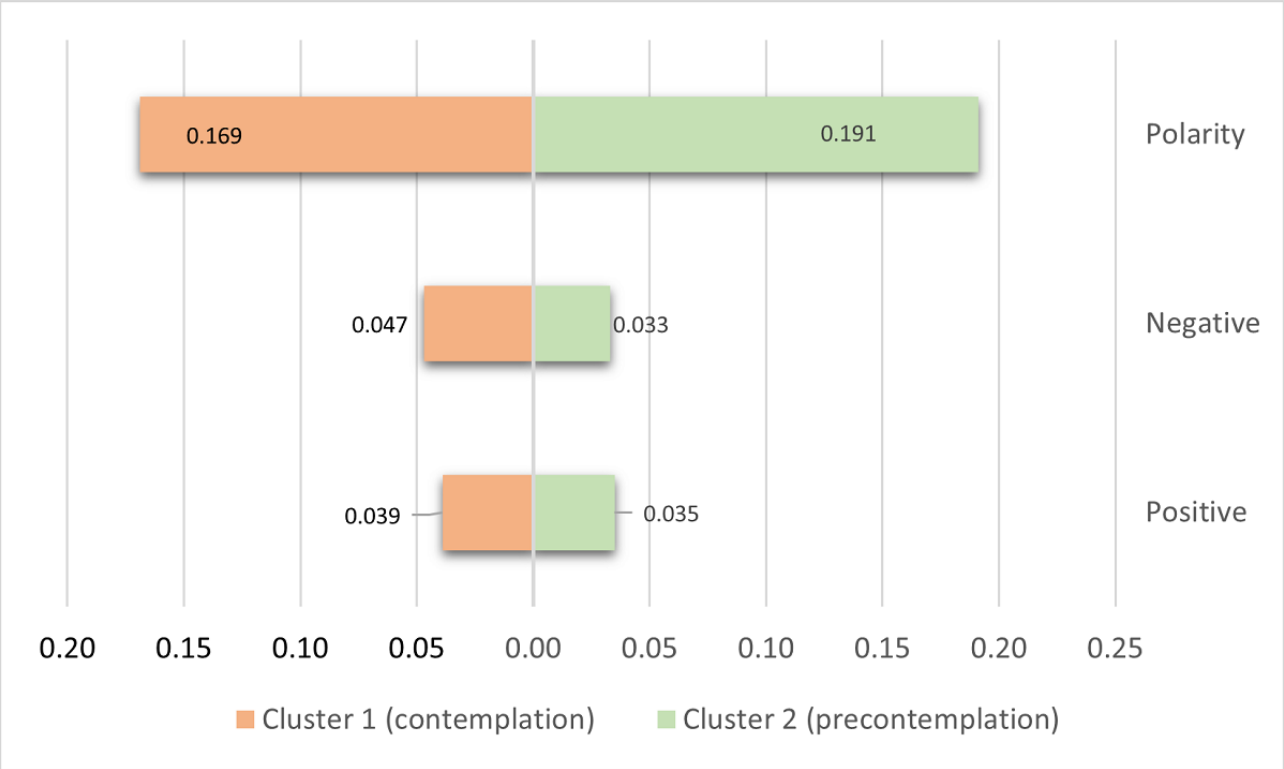


Table 20. Median values of the features considered for the contemplation and precontemplation clusters.

Emotion	Cluster 1 (contemplation), median	Cluster 2 (precontemplation), median
Joy	15.92	13.85
Sadness	32.88	21.95
Anticipation	17.08	14.13
Surprise	11.49	8.72
Trust	20.36	18.30
Disgust	22.22	13.92
Anger	19.44	13.72
Fear	23.23	14.91

We can observe that the results obtained for this task are consistent with the characteristics of people going through each stage; however, there is no annotated information for the validation of the classification performed, and this is an aspect that needs to be addressed in the future. Ours is the first approach to address this classification task, which is relevant given that users at the contemplation stage are more prone to seek help, and therefore, different ways to approach each type of user may be considered.

XSL•FO
RenderX

Discussion

Principal Findings

This study analyzed the elements that characterize Spanish-speaking users with AN during the stages of their recovery process. We have considered multiple points of view, including linguistic, psychological, relational, behavioral, demographical, and visual aspects, for our analysis. We believe that this is the first approach to analyze both users in treatment and those fully recovered in multiple perspectives. Furthermore, we identified the topics of interest of different types of users that used AN-related vocabulary, along with those of their followees. In this section, we report the main findings according to the RQs addressed.

With regard to RQ1, we found multiple elements that characterize and distinguish users with AN at the early, treatment, and full recovery stages. AN users tweeted more frequently at night and on weekends than focused control users. These results are consistent with clinical findings that suggest that patients with AN often report poor sleep quality and reduced sleep time [67]. In addition, the image results indicate that the analysis of visual elements is relevant for the detection of AN cases and focused control cases. In particular, our findings showed that the features extracted from the content generated by users are the most relevant for characterizing AN users, especially those related to linguistic and psychological factors, including terms that describe risk factors and the signs and symptoms of AN (ie, anorexia-related vocabulary).

With regard to RQ4, we determined the linguistic attributes that characterize Spanish-speaking users with AN and found that similar to a related work on English texts [6], the high use of first-person singular pronouns and verbs conjugated with these pronouns distinguishes AN users from control and recovered cases (P<.001). We have also observed that the AN group is characterized by a significantly lower use of articles and higher use of impersonal pronouns than control users (P<.001), which, to the best of our knowledge, has not been reported in any related work in English.

Our findings also reinforced the relevance of textual elements through the development of a deep learning model for the detection of AN-related tweets, which was constructed to address RQ5. We explored the change in the ratio of posts related to AN tweeted by users at each stage and found that users at the early stages of AN posted more AN-related tweets. In addition, highly significant differences among AN cases, recovered cases, and control cases (P<.001) were observed. There were also very significant differences between the AN and treatment groups (P=.004) in terms of this feature, that is, the proportion of tweets related to AN significantly changed depending on the recovery stage, indicating the progress in the recovery process of social media users with AN.

Among the relational factors explored (RQ2), we found that the AN and focused control groups could be identified by analyzing the structure of their social network (clustering approach). Among focused control users, there were several organizations and specialists for the treatment and prevention of EDs. The high polarization noted among the AN- and focused control–related communities reinforces the findings of previous studies conducted on networks of English speakers [17,18], which reported limited interactions between pro-ED and prorecovery communities (RQ4). From a psychological perspective, these findings can be explained by the elements that characterize people at the precontemplation stage according to the transtheoretical model of health behavior change, where people are in denial of their unhealthy conditions and tend to feel supported by their equals (pro-ED community members) [4,5], resulting in a rejection of prorecovery content. Regarding the topics of interest of users and their followees (RQ3), we found that the interests of AN users and their followees were highly correlated (ρ=0.96). We also observed a higher correlation between the treatment followees and focused control followee groups (ρ=0.86) and the recovered followees versus focused control followees groups (ρ=0.86) in comparison with the AN followees versus focused control followees groups (ρ=0.69). These results suggest that more interests are shared with focused control users as the recovery process advances.

With regard to RQ6, we posed a question regarding the detection of users at the precontemplation stage and users at the contemplation stage. After applying a clustering process and using emotion indicators, we identified two clusters, where one was characterized by a predominance of positive emotions over the negative ones, suggesting that users at the precontemplation stage could be the members of this cluster. On the other hand, the second cluster showed values that characterize people at the contemplation stage, where negative emotions predominate.

Finally, to address RQ7, where we explore the existence of differences among our control groups (random and focused control users), we observed that the focused control group had three times more organizations' accounts than the random control group. We also noticed that the main differences among these groups were found in linguistic attributes, especially for focused control users that were characterized by the use of a reduced number of swearing terms and more anorexia-related vocabulary terms. These findings complement our prior assumptions that focused control users were mostly organizations, specialists, and clinicians, corresponding to a prorecovery community [6,17,18].

Given our findings, another relevant aspect to discuss is the indirect and unintended role of social platforms in the promotion of harmful content among users with and without EDs. A previous study [68] reported significant decreases in caloric intake among people exposed to pro-ED websites between preexposure and postexposure. As recommender systems are designed, users with AN are likely to be recommended to follow users similar to them (other AN users), and we believe that this indirectly contributes to the reinforcement of unhealthy habits. Tools that are aware of these risks must thus be developed.

Limitations

The analyses performed in this study present certain limitations that are mainly given by the structure of the social platform analyzed, which does not provide explicit information regarding elements relevant to our analysis, such as the location, age, gender, or medical records of users. This is the main reason that

XSL•FO
RenderX

has led us to infer information based on the analysis of the users' posts; therefore, the accuracy of our results is limited to the performance of the tools we have applied. This applies to the demographical features inferred (age groups and gender); the analysis of elements that are related to the signs and symptoms of anorexia (terms related to risk factors and AN-related vocabulary); image analysis tools; and the inference of aspects that involve the location of the user, such as the weekend tweet count ratio and the sleep period tweeting ratio. This last feature is calculated in a way that overcomes the issue of not knowing the difference in the posting time according to the user's time zone. This aspect is also an issue regarding the tweeting frequency in different periods of the year, as the seasons change according to the location of users. In this sense, a Spanish variant classifier could be useful to distinguish users from different locations; however, this would still be insufficient for the cases of migrants.

We did not measure the biases introduced by the platform in our data set; however, our samples seemed to be representative of the reality in terms of age and gender despite most Twitter users being male and middle-aged [69]. The AN group users were mostly female teenagers, consistent with the age group with the highest incidence rate (15- to 19-year-old females) for AN [54]. We also considered the limitations owing to the accuracy of the translation of terms to English for the annotation and use of Empath as a topic detection tool.

We also considered the limitations that pose the characteristics of users who have a preference for Twitter as a platform and who choose to make their tweets publicly available, which might differ from those that keep their profiles private. It is important to recall that our study is limited to users who make use of Twitter; therefore, the analysis of the behavior of users from other social platforms and even of people with AN that do not have accounts on any social platforms is out of our reach.

Reproducibility and Ethical Concerns

This work was approved by the ethical review board of Pompeu Fabra University. To avoid processing and storing personal or sensitive data, a proper process of data transformation and anonymization was followed. We only stored the extracted transformed features. Our approach corresponded to an observational study, with findings that implied that the detection of AN on social media is viable; however, the further development and deployment of detection tools are required. Thus, it is necessary to perform a prior and proper risk-benefit assessment accompanied by the analysis of the legal frameworks that regulate such activities. It is also important to address the potential misusage capabilities of such tools [16,70].

Regarding the reproducibility of this work, Twitter's policies on the distribution of the data collected through its API should be respected. No information that could lead to the identification of the users included in our study will be shared, as we did not store any personal information [71]. However, the values of the features calculated are available upon reasonable request and after a proper evaluation of the use purpose.

Future Work

We believe that our findings are relevant to the development of predictive models that can assist specialists in the detection of users with AN and that can display indicators for risk factors as well as signs and symptoms that characterize AN. A proper risk assessment process should be performed before the deployment of such models.

Another aspect that we intend to analyze is the area of recommender systems. According to our findings, AN users mainly followed AN-related content, which reinforces their harmful behaviors. Thus, a recommender system that breaks this bubble and reduces the polarization between AN and focused control users can be an interesting path to follow. This is relevant considering that most recommender systems of social platforms are focused on matching users with shared interests, which in this case, might not be appropriate.

Acknowledgments

This work was supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Program (MDM-2015-0502). The authors acknowledge the funding received from the Spanish project TIN2015-65464-R (Ministerio de Asuntos Económicos y Transformación Digital/Fondo Europeo de Desarrollo Regional). The authors also acknowledge the involvement and support of members of the Fundación Instituto de Trastornos Alimentarios and the Nvidia Corporation, who donated a Tesla K40 GPU and a GTX TITAN Graphics Processing Unit (GPU) used for this research.

Conflicts of Interest

None declared.

References

1. Attia E, Walsh BT. Anorexia nervosa. Am J Psychiatry 2007 Dec;164(12):1805-1810. [doi: 10.1176/appi.ajp.2007.07071151] [Medline: 18056234]
2. Jacobi C, Hayward C, de ZM, Kraemer HC, Agras WS. Coming to terms with risk factors for eating disorders: application of risk terminology and suggestions for a general taxonomy. Psychol Bull 2004 Jan;130(1):19-65. [doi: 10.1037/0033-2909.130.1.19] [Medline: 14717649]

XSL•FO
RenderX

3. Ulfvebrand S, Birgegård A, Norring C, Högdahl L, von Hausswolff-Juhlin Y. Psychiatric comorbidity in women and men with eating disorders results from a large clinical database. Psychiatry Res 2015 Dec 15;230(2):294-299. [doi: 10.1016/j.psychres.2015.09.008] [Medline: 26416590]
4. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
5. Hasler G, Delsignore A, Milos G, Buddeberg C, Schnyder U. Application of Prochaska's transtheoretical model of change to patients with eating disorders. J Psychosom Res 2004 Jul;57(1):67-72. [doi: 10.1016/S0022-3999(03)00562-2] [Medline: 15256297]
6. De Choudhury M. Anorexia on Tumblr: a characterization study. In: Proceedings of the 5th International Conference on Digital Health 2015. 2015 Presented at: DH '15: Digital Health 2015 Conference; May 18-20, 2015; Florence Italy p. 43-50. [doi: 10.1145/2750511.2750515]
7. Losada D, Crestani F, Parapar J. Overview of eRisk - early risk prediction on the internet. In: Bellot P, Trabelsi C, Mothe J, Murtagh F, Nie JY, Soulier L, et al, editors. Experimental IR Meets Multilinguality, Multimodality, and Interaction. Switzerland: Springer; 2018:343-363.
8. Cavazos-Rehg PA, Krauss MJ, Costello SJ, Kaiser N, Cahn ES, Fitzsimmons-Craft EE, et al. "I just want to be skinny.": a content analysis of tweets expressing eating disorder symptoms. PLoS One 2019;14(1):e0207506 [FREE Full text] [doi: 10.1371/journal.pone.0207506] [Medline: 30650072]
9. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
10. Amini H, Kosseim L. Towards explainability in using deep learning for the detection of anorexia in social media. In: Natural Language Processing and Information Systems. Switzerland: Springer; 2020:225-235.
11. Ramírez-Cifuentes D, Mayans M, Freire A. Early risk detection of anorexia on social media. In: Bodrunova SS, editor. Internet Science. Switzerland: Springer; 2018:3-14.
12. Guntuku SC, Preotiuc-Pietro D, Eichstaedt JC, Ungar LH. What Twitter profile and posted images reveal about depression and anxiety. In: Proceedings of the International AAAI Conference on Web and Social Media. 2019 Presented at: Thirteenth International AAAI Conference on Web and Social Media (ICWSM-19); June 11-14, 2019; Münich, Germany p. 236-246.
13. Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Puntí J, Medina-Bravo P, Velazquez DA, et al. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. J Med Internet Res 2020 Jul 07;22(7):e17758 [FREE Full text] [doi: 10.2196/17758] [Medline: 32673256]
14. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: Pro-eating disorder socialization on Twitter. J Adolesc Health 2016 Jun;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]
15. Spinczyk D, Nabrdalik K, Rojewska K. Computer aided sentiment analysis of anorexia nervosa patients' vocabulary. Biomed Eng Online 2018 Feb 02;17(1):19 [FREE Full text] [doi: 10.1186/s12938-018-0451-2] [Medline: 29394945]
16. Chancellor S, Kalantidis Y, Pater J, De Choudhury M, Shamma D. Multimodal classification of moderated online pro-eating disorder content. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver Colorado USA p. 3213-3226. [doi: 10.1145/3025453.3025985]
17. Wang T, Brede M, Ianni A, Mentzakis E. Social interactions in online eating disorder communities: a network perspective. PLoS One 2018;13(7):e0200800 [FREE Full text] [doi: 10.1371/journal.pone.0200800] [Medline: 30059512]
18. Wang T, Brede M, Ianni A, Mentzakis E. Characterizing dynamic communication in online eating disorder communities: a multiplex network approach. Appl Netw Sci 2019 Apr 17;4(1):1-22. [doi: 10.1007/s41109-019-0125-4]
19. De Choudhury M, Counts S, Horvitz E, Hoff A. Characterizing and predicting postpartum depression from shared Facebook data. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2014 Presented at: CSCW'14: Computer Supported Cooperative Work; February 15-19, 2014; Baltimore Maryland USA p. 626-638. [doi: 10.1145/2531602.2531675]
20. Colombo GB, Burnap P, Hodorog A, Scourfield J. Analysing the connectivity and communication of suicidal users on Twitter. Comput Commun 2016 Jan 01;73(Pt B):291-300 [FREE Full text] [doi: 10.1016/j.comcom.2015.07.018] [Medline: 26973360]
21. Chancellor S, Pater J, Clear T, Gilbert E, De Choudhury M. #thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16: Computer Supported Cooperative Work and Social Computing; February 27-March 2, 2016; San Francisco California USA p. 1201-1213. [doi: 10.1145/2818048.2819963]
22. Scherr S, Arendt F, Frissen T, Oramas J. Detecting intentional self-harm on Instagram: development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. Soc Sci Comput Rev 2019 Mar 18;38(6):673-685. [doi: 10.1177/0894439319836389]
23. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1):15. [doi: 10.1140/epjds/s13688-017-0110-z]
24. Chancellor S, Mitra T, De Choudhury M. Recovery amid pro-anorexia: analysis of recovery in social media. In: Proc SIGCHI Conf Hum Factor Comput Syst. 2016 May Presented at: CHI Conference on Human Factors in Computing Systems;

XSL•FO
RenderX

2016; New York, NY, USA p. 2111-2123 URL: http://europepmc.org/abstract/MED/28840201 [doi: 10.1145/2858036.2858246]
25. López-Ubeda P, Plaza-Del-Arco F, Diaz-Galiano M, Urena-Lopez L, Martin-Valdivia MT. Detecting anorexia in Spanish tweets. In: Proceedings of Recent Advances in Natural Language Processing. 2019 Presented at: International Conference on Recent Advances in Natural Language Processing (RANLP); Sep 2–4, 2019; Varna, Bulgaria p. 655-663. [doi: 10.26615/978-954-452-056-4_077]
26. Leis A, Ronzano F, Mayer MA, Furlong LI, Sanz F. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. J Med Internet Res 2019 Jun 27;21(6):e14199 [FREE Full text] [doi: 10.2196/14199] [Medline: 31250832]
27. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 2014; Baltimore, Maryland, USA p. 51-60. [doi: 10.3115/v1/w14-3207]
28. Elrod K, Dykeman C. A corpus linguistic analysis of pro-anorexia public Tumblr posts written in Spanish. PsyArXiv 2019:A. [doi: 10.31234/osf.io/g2qye]
29. Les pagines pro ana I pro mia inunden la xarxa. Iqua - Asociación Contra La Anorexia Y Bulimia. 2011. URL: https://tinyurl.com/5dkxeen7 [accessed 2021-06-24]
30. Sample realtime tweets. Twitter. URL: https://developer.twitter.com/en/docs/tweets/sample-realtime/api-reference/get-statuses-sample [accessed 2021-06-24]
31. Leichner P. A new treatment approach to eating disorders in youth. B C Med J 2005;47(1):23-27 [FREE Full text]
32. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583. [doi: 10.2307/2280779]
33. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
34. statsmodels.stats.proportion.proportions_ztest. Statsmodels Org. URL: https://www.statsmodels.org/stable/generated/statsmodels.stats.proportion.proportions_ztest.html [accessed 2021-06-24]
35. Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1181]
36. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. J Am Med Assoc 2016 Aug 02;316(5):533-534. [doi: 10.1001/jama.2016.7653] [Medline: 27483067]
37. Ramírez-Esparza N, Pennebaker J, García FA, Suriá R. La psicología del uso de las palabras: Un programa de computadora que analiza textos en español. In: Revista Mexicana de Psicología. Distrito Federal, México: Sociedad Mexicana de Psicología A.C; 2007:85-99.
38. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
39. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
40. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. Comput Intell 2013;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]
41. Hofman E. Senti-py. GitHub, Inc. URL: https://github.com/aylliote/senti-py [accessed 2021-06-24]
42. Culbert KM, Racine SE, Klump KL. Research Review: What we have learned about the causes of eating disorders - a synthesis of sociocultural, psychological, and biological research. J Child Psychol Psychiatry 2015 Nov;56(11):1141-1164. [doi: 10.1111/jcpp.12441] [Medline: 26095891]
43. Divins M. Laxantes. Farmacia Professional. 2004. URL: https://www.elsevier.es/es-revista-farmacia-profesional-3-pdf-13064573 [accessed 2021-07-06]
44. Fast E, Chen B, Bernstein M. Empath: Understanding topic signals in large-scale text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose California USA p. 4647-4657. [doi: 10.1145/2858036.2858535]
45. Han S. googletrans 3.0.0. PyPI Library. 2020. URL: https://pypi.org/project/googletrans/ [accessed 2021-06-24]
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12(1):2825-2830.
47. Savand A. Stop-words. PyPI Library. 2018. URL: https://pypi.org/project/stop-words/ [accessed 2021-06-24]
48. Baziotis C, Pelekis N, Doulkeridis C. Datastories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017 Presented at: 11th International Workshop on Semantic Evaluation (SemEval-2017); August 2017; Vancouver, Canada p. 747-754. [doi: 10.18653/v1/s17-2126]
49. Shing H, Nair S, Zirikly A, Friedenberg M, Daumé III H, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology:

XSL•FO
RenderX

From Keyboard to Clinic. 2018 Presented at: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; June 2018; New Orleans, LA p. 25-36. [doi: 10.18653/v1/w18-0603]
50. Deriu J, Lucchi A, De LV, Severyn A, Müller S, Cieliebak M, et al. Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW '17: 26th International World Wide Web Conference; April 3-7, 2017; Perth Australia p. 1045-1052. [doi: 10.1145/3038912.3052611]
51. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 2014;9(6):e98679 [FREE Full text] [doi: 10.1371/journal.pone.0098679] [Medline: 24914678]
52. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
53. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media. 2009 Presented at: Third International AAAI Conference on Weblogs and Social Media; May 17–20, 2009; San Jose, California USA p. 361-362 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/13937
54. Hoek HW. Incidence, prevalence and mortality of anorexia nervosa and other eating disorders. Curr Opin Psychiatry 2006 Jul;19(4):389-394. [doi: 10.1097/01.yco.0000228759.95237.78] [Medline: 16721169]
55. Wang Z, Hale S, Adelani D, Grabowicz P, Hartmann T, Flöck F, et al. Demographic inference and representative population estimates from multilingual social media data. In: Proceedings of the 2019 World Wide Web Conference. 2019 Presented at: WWW '19: The Web Conference; May 13-17, 2019; San Francisco CA USA p. 2056-2067. [doi: 10.1145/3308558.3313684]
56. Silvester S, Tanbakuchi A, Müller P, Nunez-Iglesias J, Harfouche M, Klein A, OrganicIrradiation, jackwalker64, rreilink, NiklasRosenstein, Zulko, et al. Imageio/imageio v.0.9.0. Python Library. 2020. URL: https://zenodo.org/record/3931847#.YN427JgzbIU [accessed 2021-07-01]
57. Pillow 8.2.0. PyPI : Python Library. URL: https://pypi.org/project/Pillow/ [accessed 2021-06-24]
58. Computer vision with Chooch AI. Python API. URL: https://chooch.ai/api/#general-recognition-api [accessed 2021-06-24]
59. Levi G, Hassner T. Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. 2015 Presented at: ICMI '15: International Conference on Multimodal Interaction; November 9-13, 2015; Seattle Washington USA p. 503-510. [doi: 10.1145/2818346.2830587]
60. Rodriguez P, Gonzàlez J, Gonfaus J, Roca F. Integrating vision and language in social networks for identifying visual patterns of personality traits. Int J Soc Sci Humanity Stud 2019 Feb:6-12. [doi: 10.18178/ijssh.2019.v9.981]
61. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278-2324. [doi: 10.1109/5.726791]
62. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV, USA. [doi: 10.1109/cvpr.2016.90]
63. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 May 24;60(6):84-90. [doi: 10.1145/3065386]
64. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI, USA p. 1492-1500. [doi: 10.1109/cvpr.2017.634]
65. Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. arXiv: Computer Science - Machine Learning. 2016. URL: https://arxiv.org/abs/1608.03983 [accessed 2021-06-24]
66. Plutchik R. A psychoevolutionary theory of emotions. Soc Sci Inf 2017 Feb 06;21(4-5):529-553. [doi: 10.1177/053901882021004003]
67. Padez-Vieira F, Afonso P. Sleep disturbances in anorexia nervosa. Adv Eating Disorders 2016 May 04;4(2):176-188. [doi: 10.1080/21662630.2016.1175958]
68. Jett S, LaPorte DJ, Wanchisn J. Impact of exposure to pro-eating disorder websites on eating behaviour in college women. Eur Eat Disord Rev 2010;18(5):410-416. [doi: 10.1002/erv.1009] [Medline: 20572210]
69. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
70. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
71. Chancellor S, Birnbaum M, Caine E, Silenzio V, De Choudhury M. A taxonomy of ethical tensions in inferring mental health states from social media. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; January 29-31, 2019; Atlanta GA USA p. 79-88. [doi: 10.1145/3287560.3287587]

XSL•FO
RenderX

Abbreviations

AN: anorexia nervosa
API: application programming interface
BoW: bag-of-words
CNN: convolutional neural network
DTTR: day time tweets ratio
ED: eating disorder
LIWC: Linguistic Inquiry and Word Count
LR: logistic regression
pro-ED: pro-eating disorder
RQ: research question
STTR: sleep time tweets ratio

Edited by R Kukafka; submitted 21.11.20; peer-reviewed by K Usher, M Roncero, Y Hwang; comments to author 10.02.21; revised version received 10.03.21; accepted 04.05.21; published 20.07.21.

Please cite as:
Ramírez-Cifuentes D, Freire A, Baeza-Yates R, Sanz Lamora N, Álvarez A, González-Rodríguez A, Lozano Rochel M, Llobet Vives R, Velazquez DA, Gonfaus JM, Gonzàlez J
Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis
J Med Internet Res 2021;23(7):e25925
URL: https://www.jmir.org/2021/7/e25925
doi:10.2196/25925
PMID:34283033

©Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Nadia Sanz Lamora, Aida Álvarez, Alexandre González-Rodríguez, Meritxell Lozano Rochel, Roger Llobet Vives, Diego Alejandro Velazquez, Josep Maria Gonfaus, Jordi Gonzàlez. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Global Public Interests and Dynamic Trends in Osteoporosis From 2004 to 2019: Infodemiology Study

Peng Wang[1], PhD; Qing Xu[1], MPH; Rong-Rong Cao[1], PhD; Fei-Yan Deng[1], PhD; Shu-Feng Lei[1], PhD

Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University Medical College, Soochow University, Suzhou, China

Corresponding Author:
Shu-Feng Lei, PhD
Center for Genetic Epidemiology and Genomics, School of Public Health
Soochow University Medical College
Soochow University
199 Renai Road
Suzhou, 215123
China
Phone: 86 512 65883227
Fax: 86 512 65883227
Email: leisf@suda.edu.cn

Abstract

Background: With the prolonging of human life expectancy and subsequent population aging, osteoporosis (OP) has become an important public health issue.

Objective: This study aimed to understand the global public search interests and dynamic trends in "osteoporosis" using the data derived from Google Trends.

Methods: An online search was performed using the term "osteoporosis" in Google Trends from January 1, 2004, to December 31, 2019, under the category "Health." Cosinor analysis was used to test the seasonality of relative search volume (RSV) for "osteoporosis." An analysis was conducted to investigate the public search topic rising in RSV for "osteoporosis."

Results: There was a descending trend of global RSV for "osteoporosis" from January 2004 to December 2014, and a slowly increasing trend from January 2015 to December 2019. Cosinor analysis showed significant seasonal variations in global RSV for "osteoporosis" (P=.01), with a peak in March and a trough in September. In addition, similar decreasing trends of RSV for "osteoporosis" were found in Australia, New Zealand, Ireland, and Canada from January 2004 to December 2019. Cosinor test revealed significant seasonal variations in RSV for "osteoporosis" in Australia, New Zealand, Canada, Ireland, UK, and USA (all P<.001). Furthermore, public search rising topics related to "osteoporosis" included denosumab, fracture risk assessment tool, bone density, osteopenia, osteoarthritis, and risk factor.

Conclusions: Our study provided evidence about the public search interest and dynamic trends in OP using web-based data, which would be helpful for public health and policy making.

(J Med Internet Res 2021;23(7):e25422) doi:10.2196/25422

KEYWORDS

global public interest; Google trends; osteoporosis; seasonality; trends; infodemiology; information seeking; web-based information

Introduction

Overview

Osteoporosis (OP) is a generalized skeletal disorder characterized by reduction in bone mineral content, low bone mineral density, and deterioration of bone structure, all of which will eventually lead to bone fragility and an increased susceptibility to fracture [1-3]. OP is commonly described as a "silent disease" because it lacks obvious signs, and has now become an important public health issue, with estimates indicating that nearly 200 million individuals are impacted worldwide [4]. Multiple risk factors are thought to contribute to the increased likelihood of OP, including sex, age, race, hormone levels, dietary factors, and lifestyle choices [5-7].

Globally, internet has become an important platform for people to seek and share health-related information. The vast amount of new data derived from social media and search engines has shown potential values in investigating, nowcasting, and

XSL•FO
RenderX

forecasting human behaviors and diseases [8,9]. Research on internet data are usually referred to as "infodemiology studies" [10,11].

Infodemiology is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [10,11]. Infodemiology data can be collected from social media sources (Facebook, Twitter, and Instagram) and search engines (Google, Bing, and Baidu) in near real-time [12,13]. Such data allow to predict outbreaks of diseases, aid in monitoring disease syndromic surveillance as well as in detecting and quantifying disparities in health information availability [14].

Google Trends

Google Trends, as an accessible online tool, provides both real-time and archived information on Google search queries from 2004. So far, many studies have demonstrated that Google Trends could be regarded as a reliable tool for examining human behaviors, measuring the change in interest in controversial issues, analyzing public's reaction to various outbreaks or incidents, and investigating seasonal trends related to various diseases and health issues [15-20]. This new approach is also gaining importance in disease surveillance studies and could serve as an effective complement to traditional, time-consuming survey methods [21-25].

Google Trends has increasingly become a meaningful health source for both laypeople and health professionals. The web-based information on Google Trends has been recognized as a surrogate tool for estimating epidemiology and gathering data on patterns of disease and human behaviors [15,20,23]. In fact, data from internet sources could serve as a real-time surveillance tool and a supplement for health care systems, so as to allocate appropriate resources for specific moments with higher disease burden.

Google Trends, however, has not yet been used to investigate the public search interest and trends in OP. Therefore, we conducted this study to better understand the utility of Google Trends data for exploring global public search interest and dynamic trends in OP over time.

Methods

Keyword Selection

Data on internet search for "osteoporosis" were obtained from Google Trends, which provided the relative search volume (RSV) for the aforesaid search term. To make reasonable comparisons between different search terms, Google Trends adjusts the search results to the time and location of a search term by dividing each data point by the total searches of the geography and period, and then by scaling these resulting numbers based on a given search term's proportion. The higher scores represent higher RSV. The data points can be downloaded from Google Trends in ".csv" format. To avoid selection bias, Google Trends excludes all of the repeated searches from the same person during a short span of time [15].

Region and Period Selection

On April 20, 2020, the keyword "osteoporosis" was searched by individually selecting the countries "Worldwide," "Australia," "New Zealand," "Canada," "Ireland," "UK," and "USA" and choosing the category as "Health." The corresponding time-series data from January 2004 to December 2019 were then downloaded. The time-series data in our study were not a product of comparison between countries, but rather a longitudinal data on the RSV for "osteoporosis" in a single country. The period selection was representative and appropriate for our study, and contained retrospective data over the past 16 years.

Statistical Analysis

Cosinor analysis was utilized to investigate the seasonal patterns of RSV for "osteoporosis," where the RSV was regressed onto a sine and a cosine term of transformations of the time variable and represented as a sine curve that could be applied to test the seasonality [26,27]. A time-series plot was used to demonstrate the consistency in seasonal patterns. Statistical analysis on seasonality was conducted using the "season" package in R version 3.4.4 (R Foundation for Statistical Computing), while an analysis on OP-related topics was performed using Google Trends data [28]. Statistical significance was set as 2-tailed $P<.05$.

Availability of Data and Material

The data and material that support the findings of this study are available from public data sets that could be found in Google Trends.

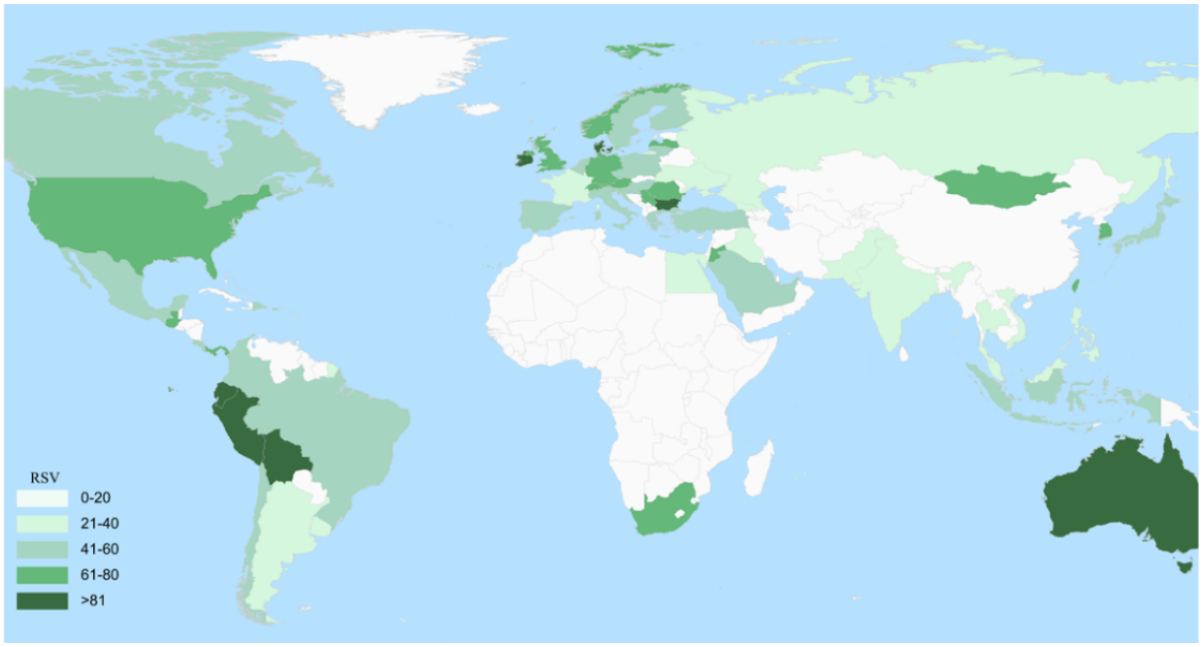
Results

Global Search Popularity for "Osteoporosis"

As displayed in Figure 1, the color intensity represents the RSV of Google searches performed during the studied period. The highest RSV for "osteoporosis" was found in Bolivia (100), Peru (95), Ireland (95), Australia (91), Bulgaria (87), Denmark (87), Singapore (73), Mongolia (73), South Africa (73), UK (69), New Zealand (69), USA (65), Panama (65), Puerto Rico (60), and Canada (60).

XSL•FO
RenderX

Figure 1. Graphic map of search popularity for "osteoporosis" by location (worldwide). RSV: relative search volume.



Global Search Trend and Seasonal Patterns for RSV of "Osteoporosis"

On a worldwide scale, there was a descending trend of RSV for "osteoporosis" from January 2004 to December 2014, and a slowly increasing trend from January 2015 to December 2019 (Figure 2A). Moreover, cosinor analysis suggested a significant seasonal pattern in RSV for "osteoporosis" (P=.01), with a peak in March and a trough in September (Figure 2B and Table 1).

Figure 2. Time series plots for the worldwide relative search volume of osteoporosis from January 01, 2004, to December 31, 2019 (A), and the plots of cosinor models for the seasonal variation in the worldwide relative search volume of osteoporosis (B).

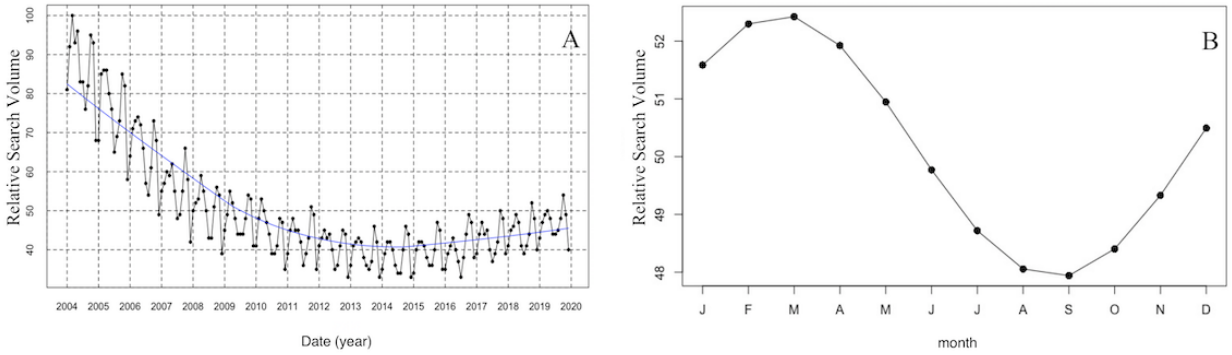


Table 1. The seasonal variations in relative search volume for "osteoporosis."

Location	Number of observations	Amplitude	Phase month	Low point month	P value[a]
Worldwide	192	2.3	2.7	8.7	.01
Australia	192	7.3	6.3	12.3	<.001
New Zealand	192	2.7	5.7	11.7	<.001
Canada	192	6.2	1.5	7.5	<.001
Ireland	192	3.6	2.0	8.0	<.001
UK	192	5.7	2.2	8.2	<.001
USA	192	4.2	1.7	7.7	<.001

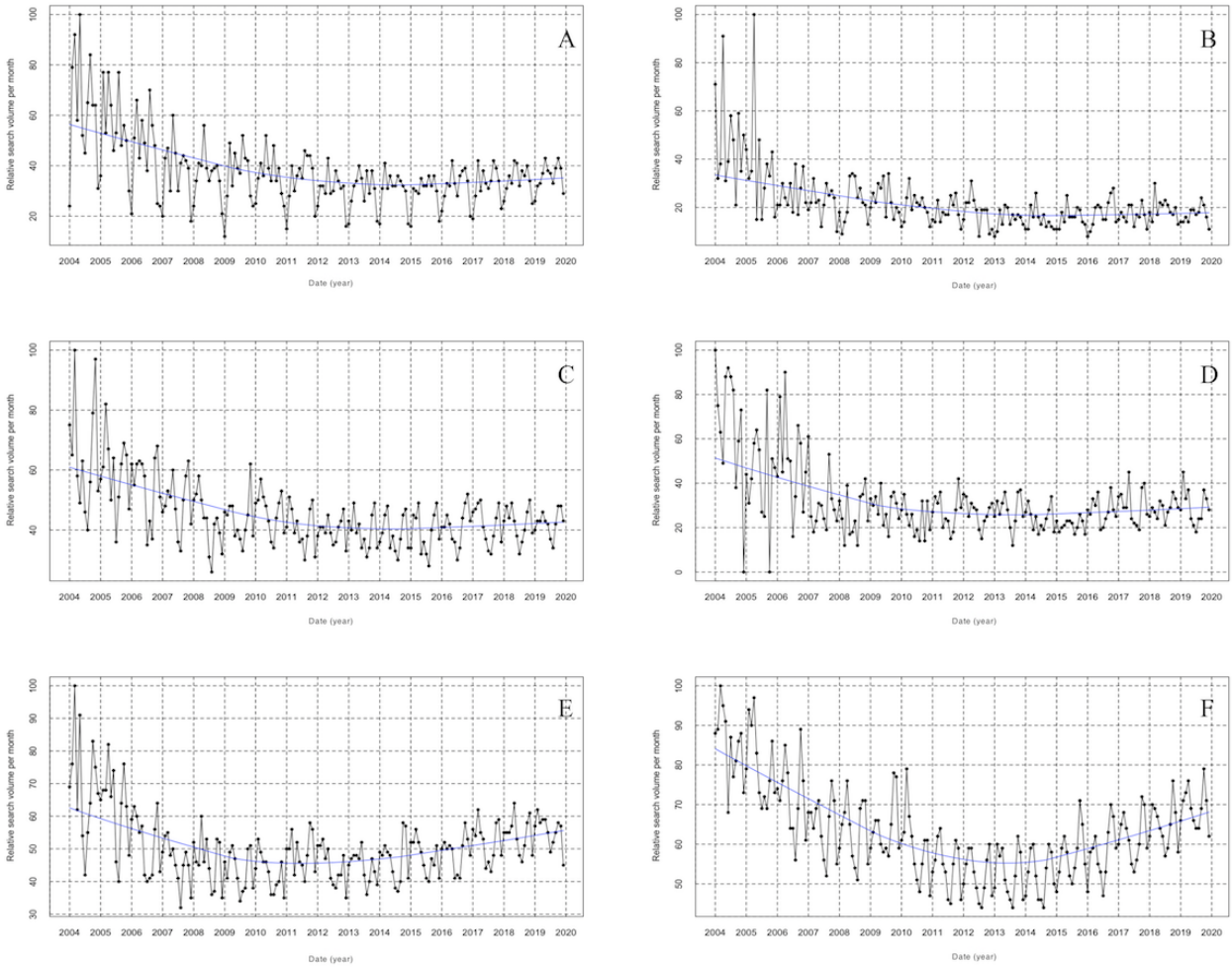
[a]Statistical significance was set as P value <.05.

XSL•FO
RenderX

Search Trend and Seasonal Pattern for RSV of "Osteoporosis" in Six English-Speaking Countries

A similar decreasing trend of RSV for "osteoporosis" was found in Australia, New Zealand, Ireland, and Canada from January 2004 to December 2019 (Figure 3A-D). In the UK, there was first a decreasing trend (from January 2004 to December 2012) and then a progressively increasing trend (from January 2013 to December 2019) of RSV for "osteoporosis" (Figure 3E). In addition, RSV for "osteoporosis" in the USA presented a descending trend from January 2004 to December 2014, and an increasing trend from January 2015 to December 2019 (Figure 3F).

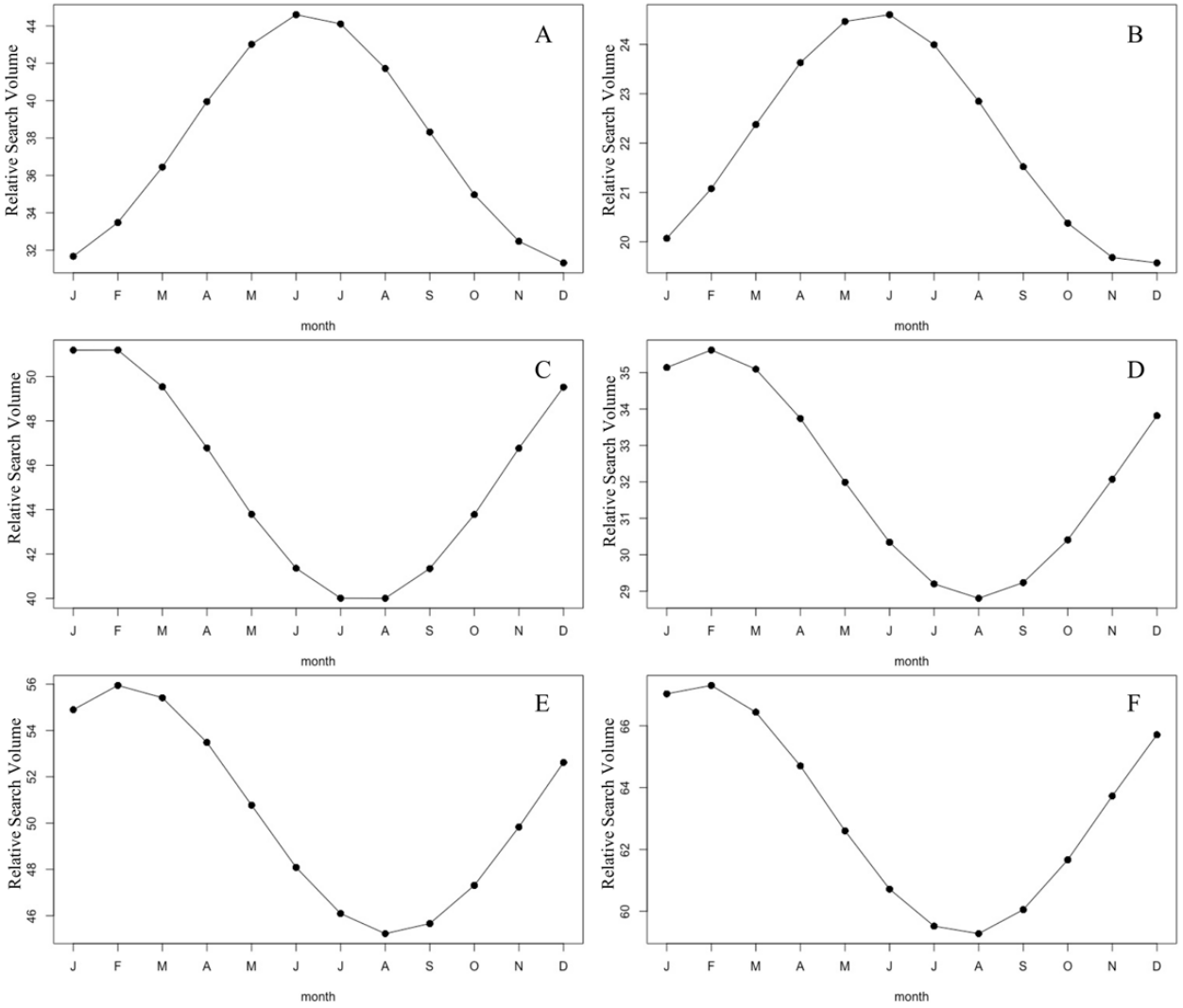
Figure 3. Time series plots for the relative search volume of osteoporosis Australia (A), New Zealand (B), Canada (C), Ireland (D), UK (E), and USA (F) from January 01, 2004, to December 31, 2019.



The results of cosinor test showed significant seasonal variations of RSV for "osteoporosis" in Australia, New Zealand, Canada, Ireland, UK, and USA (all $P<.001$; Figure 4; Table 1). RSV for "osteoporosis" in 2 southern hemisphere countries (Australia and New Zealand) peaked in winter months (June/July) and was at its lowest point in summer months (December/January; Figure 4A and B), whereas in 4 northern hemisphere countries (Canada, Ireland, UK, and USA), RSV for "osteoporosis" showed a peak in winter months (January/February) and a nadir in late summer/early autumn months (July/August/September; Figure 4C-F). There was an approximate 6-month difference in RSV for "osteoporosis" between southern and northern hemisphere countries with a reversed meteorological month (Figure 4).

XSL•FO
RenderX

Figure 4. The plots of cosinor models for the seasonal variation in the relative search volume of osteoporosis in Australia (A), New Zealand (B), Canada (C), Ireland (D), UK (E), and USA (F) from January 01, 2004, to December 31, 2019.



Rise in Public Relative Search Topics Regarding "Osteoporosis"

The rise in relative search topics was compared with the last period. The results showed that the global top rising topics were denosumab, fracture risk assessment tool (FRAX), bone tumor, hip fracture, osteomalacia, bone density, risk factor, osteoarthritis, arthritis, and osteopenia (Table 2). As for the 6 English-speaking countries, we observed that the most search rising topics were bone, bone density, osteopenia, osteoarthritis, risk factor, preventive healthcare, and the medications used for OP treatment (including denosumab, teriparatide, bisphosphonate; Table 2).

XSL•FO
RenderX

Table 2. Public relative search topics rising in "osteoporosis."

Rank	Worldwide	Australia	New Zealand	Canada	Ireland	UK	USA
	Search topics (% rising)	Search topics (% rising)	Search topics (% rising)	Search topics (% rising)	Search topics (% rising)	Search topics (% rising)	Search topics (% rising)
1	Denosumab (breakout[a])	Bone (breakout)	Bone (breakout)	Osteoarthritis (breakout)	Bone (breakout)	Bone density (breakout)	Risk factor (breakout)
2	FRAX[b] (breakout)	Calcium (breakout)	Osteoarthritis (breakout)	Guideline (breakout)	Osteoarthritis (breakout)	Osteopenia (breakout)	Denosumab (breakout)
3	Bone tumor (breakout)	Bone density (breakout)	Calcium (breakout)	Denosumab (breakout)	Irish Osteoporosis Society (breakout)	Dual-energy X-ray absorptiometry (breakout)	Teriparatide (breakout)
4	Hip fracture (breakout)	Preventive healthcare (breakout)	Vitamin D (breakout)	Bisphosphonate (breakout)	Calcium (breakout)	Osteomalacia (breakout)	Ibandronic acid (breakout)
5	Osteomalacia (1100)	Osteopenia (breakout)	Osteomalacia (breakout)	Risk factor (breakout)	Osteopenia (breakout)	FRAX (breakout)	Osteomyelitis (breakout)
6	Bone density (550)	Rheumatoid arthritis (breakout)	Bone density (breakout)	Risedronic acid (breakout)	Royal Osteoporosis Society (breakout)	Osteoarthritis (750)	Osteogenesis imperfecta (breakout)
7	Risk factor (450)	Risk factor (breakout)	Osteopenia (breakout)	Bone (650)	None	Calcium (300)	Mineral (breakout)
8	Osteoarthritis (400)	Denosumab (breakout)	Osteoporosis (40)	Osteoporosis (190)	None	Preventive healthcare (250)	Osteoarthritis (950)
9	Arthritis (250)	Osteoarthritis (300)	None	Preventive healthcare (180)	None	Osteoporosis (150)	Risk (450)
10	Osteopenia (140)	Osteoporosis (190)	None	Bone density (170)	None	Bone (60)	Bone (200)

[a]"Breakout" represents that the search term grew by more than 5000% compared with previous period.

[b]FRAX: fracture risk assessment tool.

Discussion

Principal Findings

In this study, we observed that the global internet search interest in "osteoporosis" steadily decreased from January 2004 to December 2014, whereas it slowly increased from January 2015 to December 2019. In addition, the presence of seasonal pattern in RSV for "osteoporosis" was revealed, with a peak in March and a trough in September. Moreover, we have investigated the RSV for "osteoporosis" among 6 English-speaking countries, which provided a good representation of countries in both northern and southern hemispheres. Similar change trends of RSV for "osteoporosis" were found for USA and UK as well; however, descending trends of RSV for "osteoporosis" were noted for Australia, New Zealand, Ireland, and Canada. Furthermore, the seasonal variations in RSV for "osteoporosis" among 6 English-speaking countries were confirmed, with a peak in late winter/early spring months and nadir in late summer/early autumn months. There was a nearly 6-month difference in the RSV for "osteoporosis" between southern/northern hemisphere countries with a reversed meteorological month, suggesting the presence of seasonal variations rather than calendar-driven patterns. The dynamic trends in "osteoporosis" could provide insights into the epidemiology of OP, as well as help care professionals and policy makers anticipate and prepare for this disease.

Rise in relative search topics on OP was also analyzed, with "denosumab" and "FRAX" identified as the top 2 global rising topics. Among 6 English-speaking countries, "bone" and "medications used for osteoporosis treatment" represented 2 of the most searched topics. Denosumab, also called receptor activator of nuclear factor-kappa B (RANK) ligand inhibitor, is a human monoclonal antibody used to increase bone mass and strength in the treatment of OP. The similar public search topic rising of "medications used for osteoporosis treatment" worldwide and in 6 English-speaking countries reflects the increasing public awareness of the treatment for OP other than the disease itself. FRAX is a diagnostic tool used to evaluate the probability of incurring an osteoporotic fracture. This topic rising may imply the public concerns about the possible osteoporotic fracture risk. The search topic rising interests in OP are of great importance for doctors and nurses, as they can capture the fluctuations of fast-growing topics of patients with OP and provide timely health promotion and education.

A number of studies have investigated the seasonal variation in OP presentation, and suggested that the seasonal variation in vitamin D concentration may be relevant to the occurrence of OP [29-32]. In a Greek cohort of 596 postmenopausal women

XSL•FO
RenderX

with OP, there was a seasonal variation in serum levels of 25-hydroxy vitamin D [25(OH)D], with the highest and lowest 25(OH)D levels noted in late summer/early autumn months (August/September/October) and late winter/early spring months (March), respectively. Klenk et al [33] have also demonstrated a seasonal effect on the serum 25(OH)D levels in southern Germany, with the minimum 25(OH)D serum level noted in March and the maximum in August. Another study in a Romanian population [34] also supported the association of seasonal variation with serum 25(OH)D level (highest in September and lowest in March) regardless of study subgroups. It has been demonstrated that vitamin D deficiency impairs bone mineralization and increases bone turnover, thus accelerating bone loss [35,36]. Several studies have also reported that humans in colder regions showed low cortical thickness and bone mineral density, which result in accelerated bone loss with aging [37-39]. The dynamic change of serum 25(OH)D levels may thus have an effect on the bone mass and bone architecture and play a key role in the seasonal variation in OP.

Limitations

Nevertheless, this study has several limitations that need to be acknowledged. First, Google Trends data did not measure the prevalence, but rather contained the RSV that might be influenced by several confounders. Second, the presence of already known facts and consensus could affect people's preference when searching the related term of interest. Furthermore, the influence of politics and media was evident in the trends of search volumes, which may result in sampling bias.

Despite these limitations, our study also has several strengths. The study included a large and exhaustive amount of data with a long time span, thus making the results more representative and reliable as compared with a cross-sectional study. Furthermore, the findings of this study are helpful for public health officials to facilitate aid and optimize positive health outcomes by providing resources at the best time for intervention, especially when a majority of people with health-related information needs concerning OP are engaged in the process of information seeking.

Conclusions

Overall, this study revealed a slow global increase of internet search for "osteoporosis" in recent years, and also showed a significant seasonal variation in global RSV for "osteoporosis." In addition, the presence of seasonal patterns in RSV for "osteoporosis" was found in 6 English-speaking countries. Public relative search rising topics regarding "osteoporosis" indicated the major public concerns about this disease. This study also provided evidence about the search interest of public and dynamic trends in OP through an internet search, which could provide an initial contact point for patients experiencing symptoms, and may potentially be used to expedite necessary medical evaluation. In addition, as compared with a traditional epidemiological study, web-based data could be used as a supplement to the traditional surveillance data for the early control and prevention of this disease.

Acknowledgments

We thank all the people who participated in this study.

Authors' Contributions

PW and SFL conceived the presented idea. PW and QX developed the theory and performed the computations. RRC and FYD verified the analytical methods. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

References

1. Rizzoli R. Postmenopausal osteoporosis: Assessment and management. Best Pract Res Clin Endocrinol Metab 2018 Oct;32(5):739-757. [doi: 10.1016/j.beem.2018.09.005] [Medline: 30449552]
2. Bragg S, Bain J, Ramsetty A. Endocrine Conditions in Older Adults: Osteoporosis. FP Essent 2018 Nov;474:11-19. [Medline: 30427648]
3. Torrente-Segarra V, Roig-Vilaseca D. Denosumab mid-term densitometric gain in postmenopausal osteoporosis women in clinical practice: comment on "Variability of denosumab densitometric response in postmenopausal osteoporosis". Rheumatol Int 2018 Jun;38(6):1161-1162. [doi: 10.1007/s00296-018-4007-3] [Medline: 29546636]
4. Atik OS, Gunal I, Korkusuz F. Burden of osteoporosis. Clin Orthop Relat Res 2006 Feb;443:19-24. [doi: 10.1097/01.blo.0000200248.34876.fe] [Medline: 16462421]
5. Lane NE. Epidemiology, etiology, and diagnosis of osteoporosis. Am J Obstet Gynecol 2006 Feb;194(2 Suppl):S3-11. [doi: 10.1016/j.ajog.2005.08.047] [Medline: 16448873]
6. Gonçalves CF, Meng Q. Timing metabolism in cartilage and bone: links between circadian clocks and tissue homeostasis. J Endocrinol 2019 Aug 01;243(3):R29-R46. [doi: 10.1530/JOE-19-0256] [Medline: 31470415]

XSL•FO
RenderX

7. Farman HH, Gustafsson KL, Henning P, Grahnemo L, Lionikaite V, Movérare-Skrtic S, et al. Membrane estrogen receptor α is essential for estrogen signaling in the male skeleton. J Endocrinol 2018 Dec 01;239(3):303-312. [doi: 10.1530/JOE-18-0406] [Medline: 30400010]
8. Park TH, Kim WI, Park S, Ahn J, Cho MK, Kim S. Public Interest in Acne on the Internet: Comparison of Search Information From Google Trends and Naver. J Med Internet Res 2020 Oct 26;22(10):e19427 [FREE Full text] [doi: 10.2196/19427] [Medline: 33104003]
9. Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E. Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study. J Med Internet Res 2020 Jul 24;22(7):e17633 [FREE Full text] [doi: 10.2196/17633] [Medline: 32706682]
10. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
11. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
12. Mavragani A. Infodemiology and Infoveillance: Scoping Review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
13. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
14. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
15. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
16. Bragazzi NL, Barberis I, Rosselli R, Gianfredi V, Nucci D, Moretti M, et al. How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. Hum Vaccin Immunother 2017 Feb;13(2):464-469. [doi: 10.1080/21645515.2017.1264742] [Medline: 27983896]
17. Hossain L, Kam D, Kong F, Wigand RT, Bossomaier T. Social media in Ebola outbreak. Epidemiol Infect 2016 Jul;144(10):2136-2143. [doi: 10.1017/S095026881600039X] [Medline: 26939535]
18. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating Smart Health in the US Health Care System: Infodemiology Study of Asthma Monitoring in the Google Era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24 [FREE Full text] [doi: 10.2196/publichealth.8726] [Medline: 29530839]
19. Phelan N, Kelly JC, Moore DP, Kenny P. The effect of the metal-on-metal hip controversy on Internet search activity. Eur J Orthop Surg Traumatol 2014 Oct;24(7):1203-1210. [doi: 10.1007/s00590-013-1399-3] [Medline: 24390041]
20. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]
21. Yang AC, Huang NE, Peng C, Tsai S. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. PLoS One 2010 Oct 28;5(10):e13728 [FREE Full text] [doi: 10.1371/journal.pone.0013728] [Medline: 21060851]
22. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
23. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]
24. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
25. Chang Y, Chiang W, Wang W, Lin C, Hung L, Tsai Y, et al. Google Trends-based non-English language query data and epidemic diseases: a cross-sectional study of the popular search behaviour in Taiwan. BMJ Open 2020 Jul 05;10(7):e034156 [FREE Full text] [doi: 10.1136/bmjopen-2019-034156] [Medline: 32624467]
26. Kardeş S. Seasonal variation in the internet searches for gout: an ecological study. Clin Rheumatol 2019 Mar;38(3):769-775. [doi: 10.1007/s10067-018-4345-2] [Medline: 30374747]
27. Barnett AG, Dobson AJ. Analysing Seasonal Health Data. Berlin, Germany: Springer; 2010.
28. Barnett A, Baker P, Dobson A. Analysing Seasonal Data. The R Journal 2012;4(1):5-10. [doi: 10.32614/rj-2012-001]
29. Aw D, Thain J, Ali A, Aung T, Chua WM, Sahota O, et al. Predicting fracture risk in osteoporosis: the use of fracture prediction tools in an osteoporosis clinic population. Postgrad Med J 2016 May;92(1087):267-270. [doi: 10.1136/postgradmedj-2015-133454] [Medline: 26792635]

XSL•FO
RenderX

30. Andersen R, Mølgaard C, Skovgaard LT, Brot C, Cashman KD, Chabros E, et al. Teenage girls and elderly women living in northern Europe have low winter vitamin D status. Eur J Clin Nutr 2005 Apr;59(4):533-541 [FREE Full text] [doi: 10.1038/sj.ejcn.1602108] [Medline: 15714215]
31. van der Wielen RP, Löwik MR, van den Berg H, de Groot LC, Haller J, Moreiras O, et al. Serum vitamin D concentrations among elderly people in Europe. Lancet 1995 Jul 22;346(8969):207-210. [doi: 10.1016/s0140-6736(95)91266-5] [Medline: 7616799]
32. Hansen L, Tjønneland A, Køster B, Brot C, Andersen R, Cohen AS, et al. Vitamin D Status and Seasonal Variation among Danish Children and Adults: A Descriptive Study. Nutrients 2018 Nov 20;10(11):1801 [FREE Full text] [doi: 10.3390/nu10111801] [Medline: 30463277]
33. Klenk J, Rapp K, Denkinger MD, Nagel G, Nikolaus T, Peter R, et al. Seasonality of vitamin D status in older people in Southern Germany: implications for assessment. Age Ageing 2013 May;42(3):404-408. [doi: 10.1093/ageing/aft042] [Medline: 23542723]
34. Niculescu DA, Capatina CAM, Dusceac R, Caragheorgheopol A, Ghemigian A, Poiana C. Seasonal variation of serum vitamin D levels in Romania. Arch Osteoporos 2017 Dec 11;12(1):113. [doi: 10.1007/s11657-017-0407-3] [Medline: 29230557]
35. Looker AC. Serum 25-hydroxyvitamin D and risk of major osteoporotic fractures in older U.S. adults. J Bone Miner Res 2013 May;28(5):997-1006 [FREE Full text] [doi: 10.1002/jbmr.1828] [Medline: 23184640]
36. Sai AJ, Walters RW, Fang X, Gallagher JC. Relationship between vitamin D, parathyroid hormone, and bone health. J Clin Endocrinol Metab 2011 Mar;96(3):E436-E446 [FREE Full text] [doi: 10.1210/jc.2010-1886] [Medline: 21159838]
37. Leslie WD, Weiler HA, Lix LM, Nyomba BLG. Body composition and bone density in Canadian White and Aboriginal women: the First Nations Bone Health Study. Bone 2008 May;42(5):990-995. [doi: 10.1016/j.bone.2008.01.018] [Medline: 18329354]
38. Leslie WD, Metge CJ, Weiler HA, Doupe M, Wood Steiman P, O'Neil JD. Bone density and bone area in Canadian Aboriginal women: the First Nations Bone Health Study. Osteoporos Int 2006 Dec;17(12):1755-1762. [doi: 10.1007/s00198-006-0184-0] [Medline: 16960648]
39. Wallace IJ, Nesbitt A, Mongle C, Gould ES, Grine FE. Age-related variation in limb bone diaphyseal structure among Inuit foragers from Point Hope, northern Alaska. Arch Osteoporos 2014;9:202. [doi: 10.1007/s11657-014-0202-3] [Medline: 25491658]

Abbreviations

25(OH)D: 25-hydroxy vitamin D
FRAX: fracture risk assessment tool
OP: osteoporosis
RSV: relative search volume

Edited by R Kukafka; submitted 01.11.20; peer-reviewed by A Mavragani, Z Xu; comments to author 26.11.20; revised version received 22.12.20; accepted 27.04.21; published 05.07.21.

Please cite as:
Wang P, Xu Q, Cao RR, Deng FY, Lei SF
Global Public Interests and Dynamic Trends in Osteoporosis From 2004 to 2019: Infodemiology Study
J Med Internet Res 2021;23(7):e25422
URL: https://www.jmir.org/2021/7/e25422
doi:10.2196/25422
PMID:36260400

©Peng Wang, Qing Xu, Rong-Rong Cao, Fei-Yan Deng, Shu-Feng Lei. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Search Trends to Analyze Web-Based Interest in Lower Urinary Tract Symptoms-Related Inquiries, Diagnoses, and Treatments in Mainland China: Infodemiology Study of Baidu Index Data

Shanzun Wei[1], MD; Ming Ma[1], MD; Changjing Wu[2], PhD; Botao Yu[1], MD; Lisha Jiang[3], MA; Xi Wen[2], MSc; Fudong Fu[2], BA; Ming Shi[1], MD

[1]Department of Urology, West China Hospital, Sichuan University, Chengdu, China

[2]Andrology Laboratory, West China Hospital, Sichuan University, Chengdu, China

[3]Day Surgery Center, West China Hospital, Sichuan University, China, China

Corresponding Author:
Ming Shi, MD
Department of Urology
West China Hospital
Sichuan University
37 Guoxue Alley
Chengdu, 611130
China
Phone: 86 18080173954
Email: 18080173954@189.cn

Abstract

Background: Lower urinary tract symptoms (LUTS) are one of the most commonly described urination disorders worldwide. Previous investigations have focused predominantly on the prospective identification of cases that meet the researchers' criteria; thus, the genuine demands regarding LUTS from patients and related issues may be neglected.

Objective: We aimed to examine web-based search trends and behaviors related to LUTS on a national and regional scale by using the dominant, major search engine in mainland China.

Methods: Baidu Index was queried by using LUTS-related terms for the period of January 2011 to September 2020. The search volume for each term was recorded to analyze search trends and demographic distributions. For user interest, user demand graph data and trend data were collected and analyzed.

Results: Of the 13 LUTS domains, 11 domains are available in the Baidu Index database. The Baidu search index for each LUTS domain varied from 37.78% to 1.47%. The search trends for *urinary frequency* (2011-2018: annual percent change APC=7.82%; *P*<.001), *incomplete emptying* (2011-2014: APC=17.74%; *P*<.001), *nocturia* (2011-2018: APC=11.54%; *P*<.001), *dysuria* (2017-2020: APC=20.77%; *P*<.001), and *incontinence* (2011-2016: APC=13.39%; *P*<.001) exhibited fluctuations over time. The search index trends for *weak stream* (2011-2017: APC=−4.68%; *P*<.001; 2017-2020: APC=9.32%; *P*=.23), *split stream* (2011-2013: APC=9.50%; *P*=.44; 2013-2020: APC=2.05%; *P*=.71), *urgency* (2011-2018: APC=−2.63%; *P*=.03; 2018-2020: APC=8.58%; *P*=.19), and *nocturnal enuresis* (2011-2018: APC=−3.20%; *P*=.001; 2018-2020: APC=−4.21%; *P*=.04) remained relatively stable and consistent. The age distribution of the population for all LUTS-related inquiries showed that individuals aged 20 to 40 years made 73.86% (49,218,123/66,635,247) of the total search inquiries. Further, individuals aged 40 to 49 years made 12.29% (8,193,922/66,635,247) of the total search inquiries for all LUTS-related terms. People from the east part of China made 67.79% (45,172,031/66,635,247) of the total search queries. Additionally, most of the searches for LUTS-related terms were related to those for urinary diseases to varying degrees.

Conclusions: Web-based interest in LUTS-related terms fluctuated wildly and was reflected timely by Baidu Index in mainland China. The web-based search popularity of each LUTS-related term varied significantly and differed based on personal interests, the population's concerns, regional variations, and gender. These data can be used by care providers to track the prevalence of LUTS and the population's interests, guide the establishment of disease-specific health care policies, and optimize physician-patient health care sessions.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e27029) doi:10.2196/27029

KEYWORDS

lower urinary tract symptoms; patient education; Baidu Index; infodemiology; public interest; urinary tract disorders; infoveillance; web-based search; search engines; health care policy; digital health

Introduction

Lower urinary tract symptoms (LUTS) are one of the most commonly described urination disorders worldwide [1]. There are a wide range of characteristic LUTS. These symptoms, such as urinary frequency, nocturia, urinary urgency, weak stream, hesitancy, terminal dribble, incomplete emptying, and urinary incontinence, can be categorized into three groups—storage, voiding, and postmicturition symptoms. These symptoms are generally manifestations of detrusor overactivity, sphincteric weakness, sensory bladder disorders, and prostate hyperplasia [1]. Due to the encompassing symptoms associated with sexual dysfunction and constant discomfort, LUTS significantly impair one's quality of life, social functioning, and workplace productivity and result in high health care costs [2-4]. The reported prevalence of LUTS is susceptible to numerous factors, such as diagnostic and assessment criteria (eg, International Continence Society [ICS] and The Expanded Prostate Cancer Index Composite study), evaluation tools (eg, International Prostate Symptom Score [IPSS] and Overactive Bladder Symptom Score), data collection, and various definitions, and therefore varies widely [5]. According to data from different multicenter surveys, the incidence of LUTS is between 26% and 86% [6-8]. In recent years, with the gradually heightened requirements for individuals' quality of life and the incremental attention to the early symptoms of diseases, the public awareness of problems concerning LUTS has increased [9]. Similarly, the inconsistencies in previous LUTS prevalence investigations are noteworthy in China. Although these surveys were nationally conducted, the results can only provide the incidence of LUTS in specific populations—individuals aged above 18 years [10], males aged over 40 years [11], patients with benign prostatic hyperplasia (BPH) [12], and adult Chinese women [13]—for a sample size ranging from 3023 to 18,992. In some investigations, the annual LUTS incidents presented a growing trend, but patients' recognition and health care help-seeking rates were low [10-12]. It should be noted that previous investigations have focused predominantly on the prospective identification of cases that meet the researchers' criteria. The data were collected with standardized questionnaires that were distributed via email, the agency's website, or on-site inquires [10,12]. Coupled with the lack of national surveillance and corresponding epidemiological reporting systems, the genuine demands regarding LUTS from patients and related issues may thus be neglected.

Nowadays, lives have been extensively changed by internet development. Internet searches have become peoples' first-choice method for seeking information regarding health issues [14]. The advanced, mobile cyber technology used in emerging social media search engines have enabled the internet to become more accessible to people with all kinds of quests [15]. Additionally, artificial intelligence and big data technology enable search services to provide health-related and better suited information, especially to those who feel sick or suspect that they have early symptoms. It is believed that patients have been increasingly conducting web searches for health information prior to seeing a doctor. Hence, using data from Google has been successfully practiced when reporting the incidence of rhinitis [16], surveilling the prevalence of burn injuries [17], tracking e-cigarette–related lung injury cases [18], and forecasting and analyzing the public awareness of pandemic outbreaks [19].

In mainland China, Baidu has monopolized search services. After Google was forced to shut down its services in China, 92.1% of search volume data are now found in Baidu's platform, and the usage of Baidu's platform accounts for 93.1% of search service usage [20,21]. As the top-ranked search site [22], Baidu's big data analyzing platform, Baidu Index, can reflect the genuine needs of the real world geospatially, temporally, and conclusively based on users' specified terms [23-25]. It has been proven that Baidu Index is useful for surveilling and forecasting epidemic prevalence [23]. Additionally, the user behavior portrait may provide valuable insights into the health care concerns of populations and help with examining patients' experience sentiments [24]. These results will guide the tracking of common interests and the conduction of disease prevention and control education in a more focused manner [25].

In September 2020, the netizen population size reached 940 million in China [26]. For the first time, the China Internet Network Information Center has promoted the usage of internet health care services due to their better privacy and readiness. With 766 million users using the Baidu search service actively, its usage in relation to health inquiries and symptom confirmation have accounted for 63.16% of search service use [26,27]. Therefore, a pragmatic approach to monitoring LUTS incidents is analyzing data from the Baidu search index (BSI). The primary objective of this study was to assess the prevalence of LUTS by analyzing internet search activity and examining the validity of related topics. We also aimed to investigate the national demographics of people with LUTS and mass health-seeking concerns.

Methods

Keyword Selection and Data Retrieval

This study mainly analyzed the temporal search trends of LUTS-related terms in China. Symptoms domains were identified by referring to the ICS's reports [1,5,28-30]. The Chinese terms were selected by referring to the ICS's LUTS translation recommendations to reduce results bias resulting from different language habits. All possible synonymous and derivative keywords for each term were screened [24,31] (Multimedia Appendix 1). Additionally, the availability of each keyword was examined on the Baidu Index platform. All

XSL•FO
RenderX

available LUTS-related terms are listed in Multimedia Appendix 2.

In the trend analysis module, the search index value for each keyword is an absolute numerical value that is converted from normalized daily search counts. Hence, we obtained the search index values of each keyword from January 1, 2011, to September 12, 2020. All trend data were collected at the provincial and national levels [24,31]. For terms with multiple available keywords, we summarized the daily index value of all related keywords [25,31]. In addition to the trend data module, the keyword search-demand module shows the most noted correlating issues and sorts these issues by search frequency. The demographic portrait module records users' ages and the regional distributions of inquired keywords. Therefore, we also collected user demand graph data and geodemographic data from the Baidu database to analyze netizens' demands and the public awareness of LUTS.

Data Analysis

To analyze the search trend for each term, the corresponding BSI data were plotted sequentially to describe public attention. We calculated the medians and IQRs of daily search index values to describe annual and seasonal changes according to the normality and homogeneity of variance of the data. The definition of seasonal changes was taken from the standard definition provided by the National Meteorological Administration of China [32]. Statistical differences in index values from different periods were determined via the Mann-Whitney test. A Spearman correlation analysis was used to analyze the correlations among the search indices of common diseases with LUTS and LUTS-related terms. A P value of <.05 was considered statistically significant. The changes in trends over time for each domain was identified with the Joinpoint Regression model (Program Version 4.7.0.0; Statistical Research and Applications Branch, National Cancer Institute). This model is an ideal tool that is suited for examining big data over time and testing whether an apparent change in a trend is statistically significant. Annual percent change (APC) is the summary measure of a trend over a prespecified fixed interval [33].

Statistical Analysis

All databases were constructed with Excel 2019 (Microsoft Corporation). The Mann-Whitney test and Spearman correlation analysis were conducted using IBM SPSS, version 22.0 (IBM Corporation). We used Prism 8 for macOS, version 8.4.0 (455) (GraphPad Software Incorporated) to conduct statistical analysis and create figures.

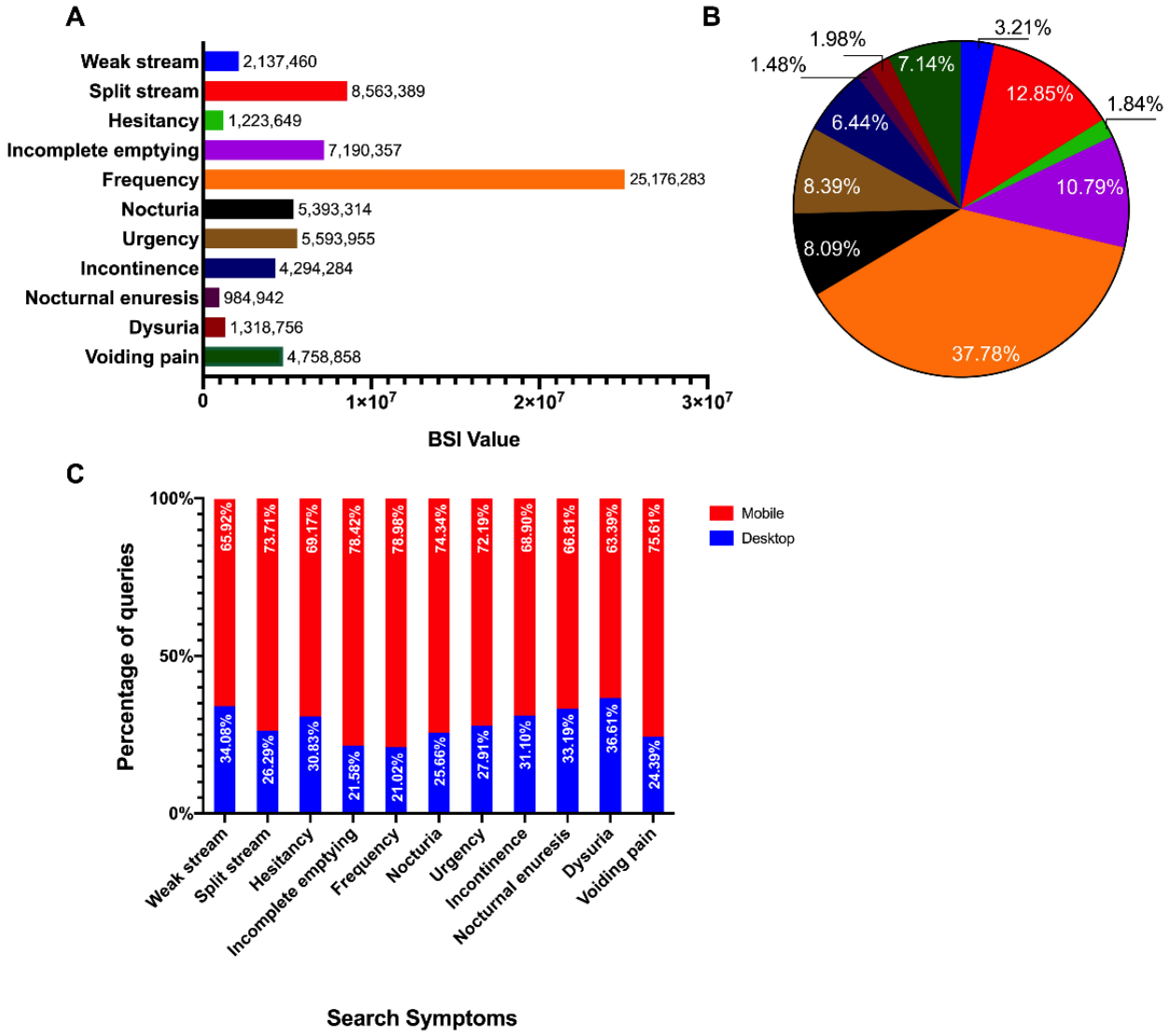
Results

Web-Based Data Trends of LUTS-Related Terms

We summarized the total BSI values of LUTS-related terms from the past 10 years. Of the 13 LUTS domains, 11 domains were available in the Baidu Index database. The BSI for each LUTS term varied greatly. The term *urinary frequency* had the highest total BSI value (37.78%), whereas the term *nocturnal enuresis* only had a total BSI value of 1.47%. Additionally, we found that 75.31% (50,183,464/66,635,247) requests were from nondesktop computers, revealing that users were more likely to conduct searches with their mobile devices (Figure 1).

XSL•FO
RenderX

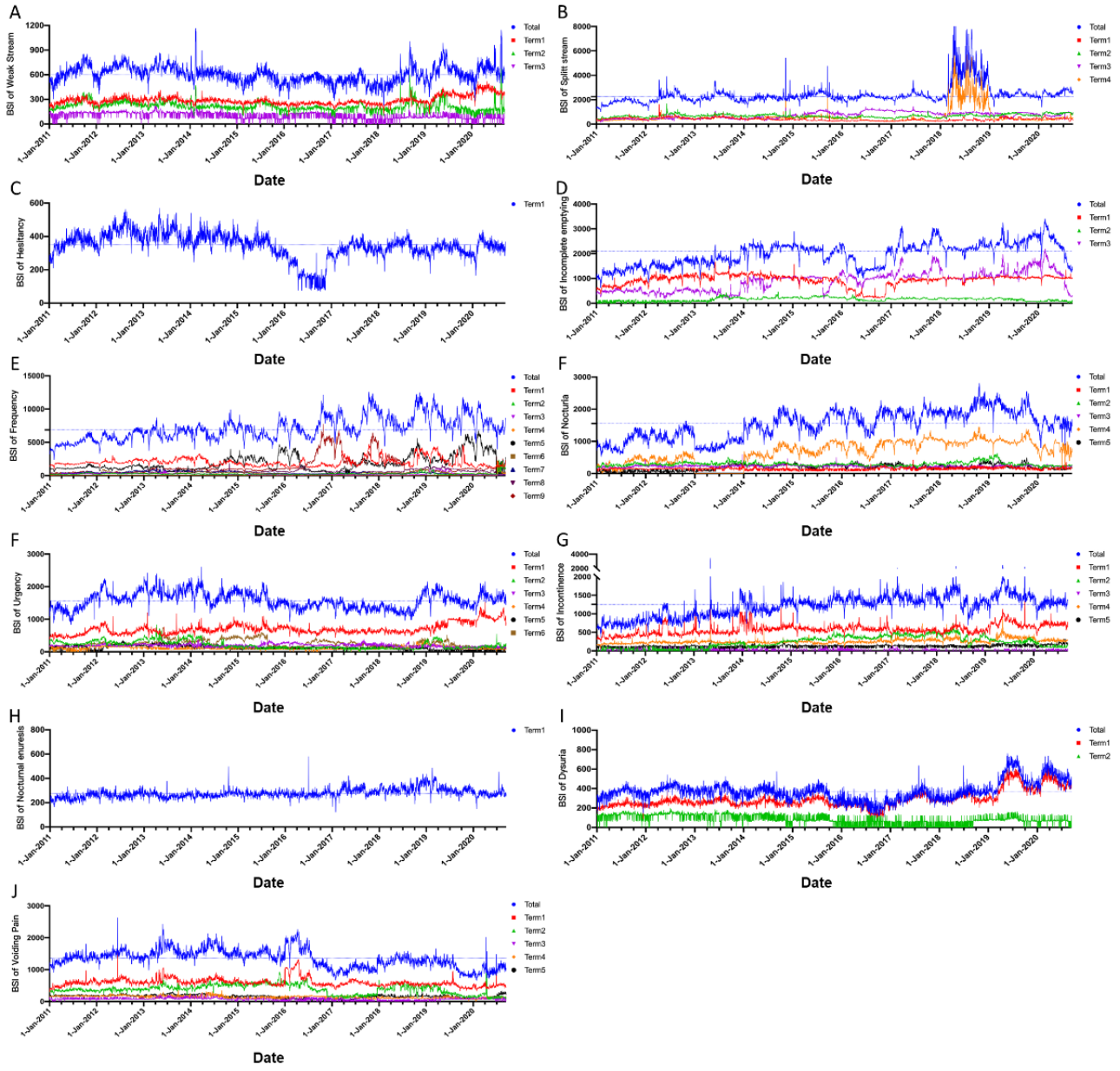
Figure 1. Web-based interest in LUTS domains. A: Total BSI value of each LUTS domain. B: BSI proportion of each LUTS domain. C: Ratio between mobile internet access and desktop access. BSI: Baidu search index; LUTS: lower urinary tract symptoms.



We created a daily time series curve for the BSIs of each LUTS-related keyword for mainland China (Figure 2) and sorted the medians of each BSI value by year and season (Multimedia Appendix 3). Based on the median annual BSI for each LUTS-related term, the search trends for *urinary frequency* (2011-2018: APC=7.82%; *P*<.001), *incomplete emptying* (2011-2014: APC=17.74%; *P*<.001), *nocturia* (2011-2018: APC=11.54%; *P*<.001), *dysuria* (2017-2020: APC=20.77%; *P*<.001), and *incontinence* (2011-2016: APC=13.39%; *P*<.001) relatively grew for a period of time. The search index trends for *weak stream* (2011-2017: APC=−4.68%; *P*<.001; 2017-2020: APC=9.32%; *P*=.23), *split stream* (2011-2013: APC=9.50%; *P*=.44; 2013-2020: APC=2.05%; *P*=.71), *urgency* (2011-2018: APC=−2.63%; *P*=.03; 2018-2020: APC=8.58%; *P*=.19), and *nocturnal enuresis* (2011-2018: APC=3.20%; *P*<.001.; 2019-2020: APC=−4.21%, *P*=.44) remained relatively stable and consistent (Multimedia Appendix 4). A notable spike in search volume was found for the term *split stream* in the year of 2018, and this mainly resulted from a surge in searches for term 4 (尿尿分叉; ie, *Wee Wee split*). With regard to the terms *hesitancy* and *voiding pain*, an annual downward trend was found for each keyword. With regard to seasonal differences, though the trends for *frequency* (*P*=.057), *incontinence* (*P*=.36), and *nocturia* (*P*=.27) seemed to fluctuate in a pattern, the seasonal BSI gap for each term was not significant.

XSL•FO
RenderX

Figure 2. Real-time daily trends of web-based interest in each lower urinary tract symptoms–term over the last 10 years. BSI: Baidu search index.

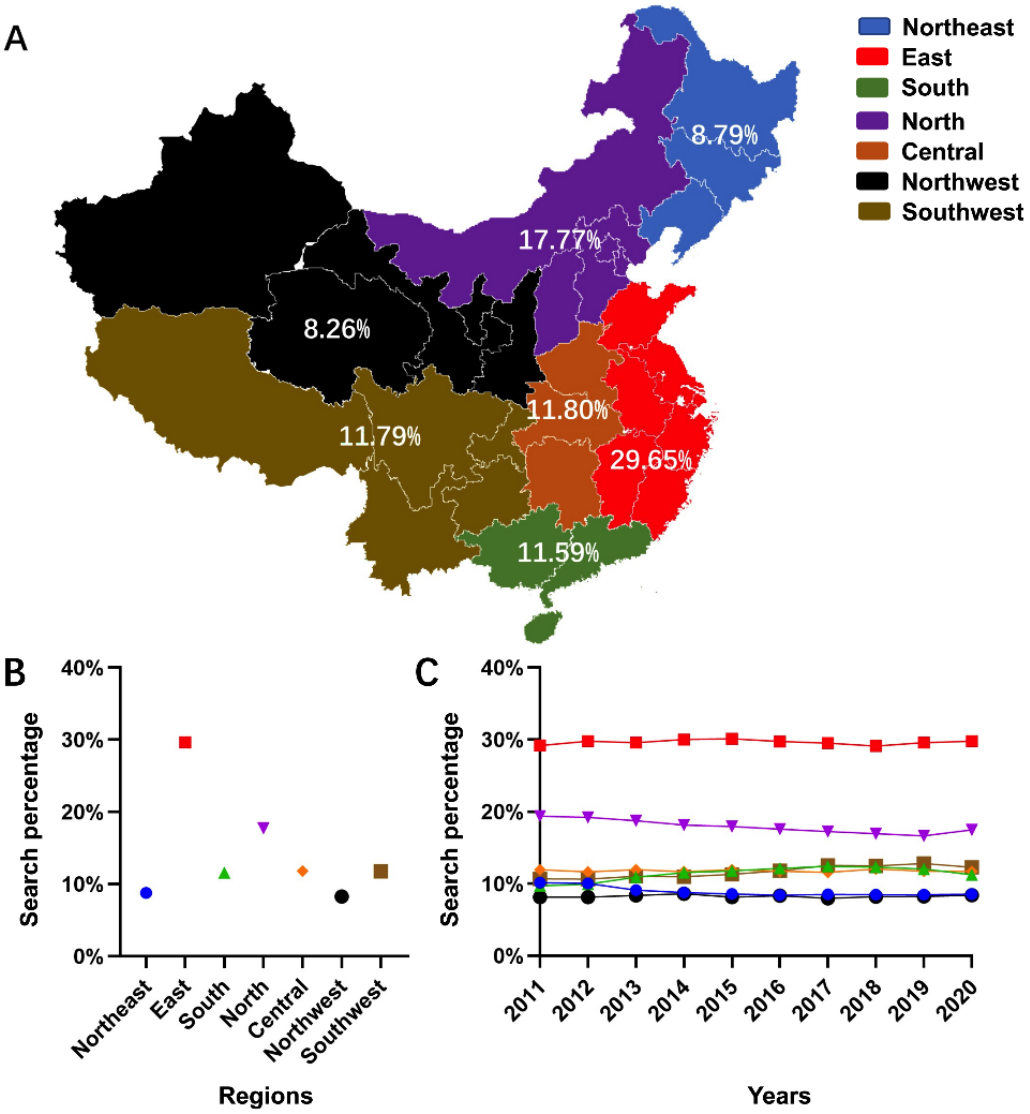


Geographic Differences

Geographic differences in LUTS-related terms' BSI data were calculated based on provincial data and sorted according to Chinese administrative divisions. These regions were Northeast China, East China, South China, North China, Central China, and Northwest and Southwest China. In Figure 3, the 10-year regional BSI proportions for all LUTS-related terms are presented in a map of mainland China. Additionally, the regional BSI proportions for each year are presented. It was notable that people from the east part of China (East, North, Northeast, and South China) made 67.79% (45,172,031/66,635,247) of the total search queries. However, the queries from West China (northwest and southwest) only accounted for 20.05% (13,360,352/66,635,247) of the search queries. The regional BSI proportion of each LUTS-related term and the annual trends are presented in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 3. Regional distribution of web-based interest based on lower urinary tract symptoms–related searches over the last 10 years (with available data). A: Regional rates for each area. B: 10-year total search rates for each area. C: Annual trend of the Baidu search indices for each region.

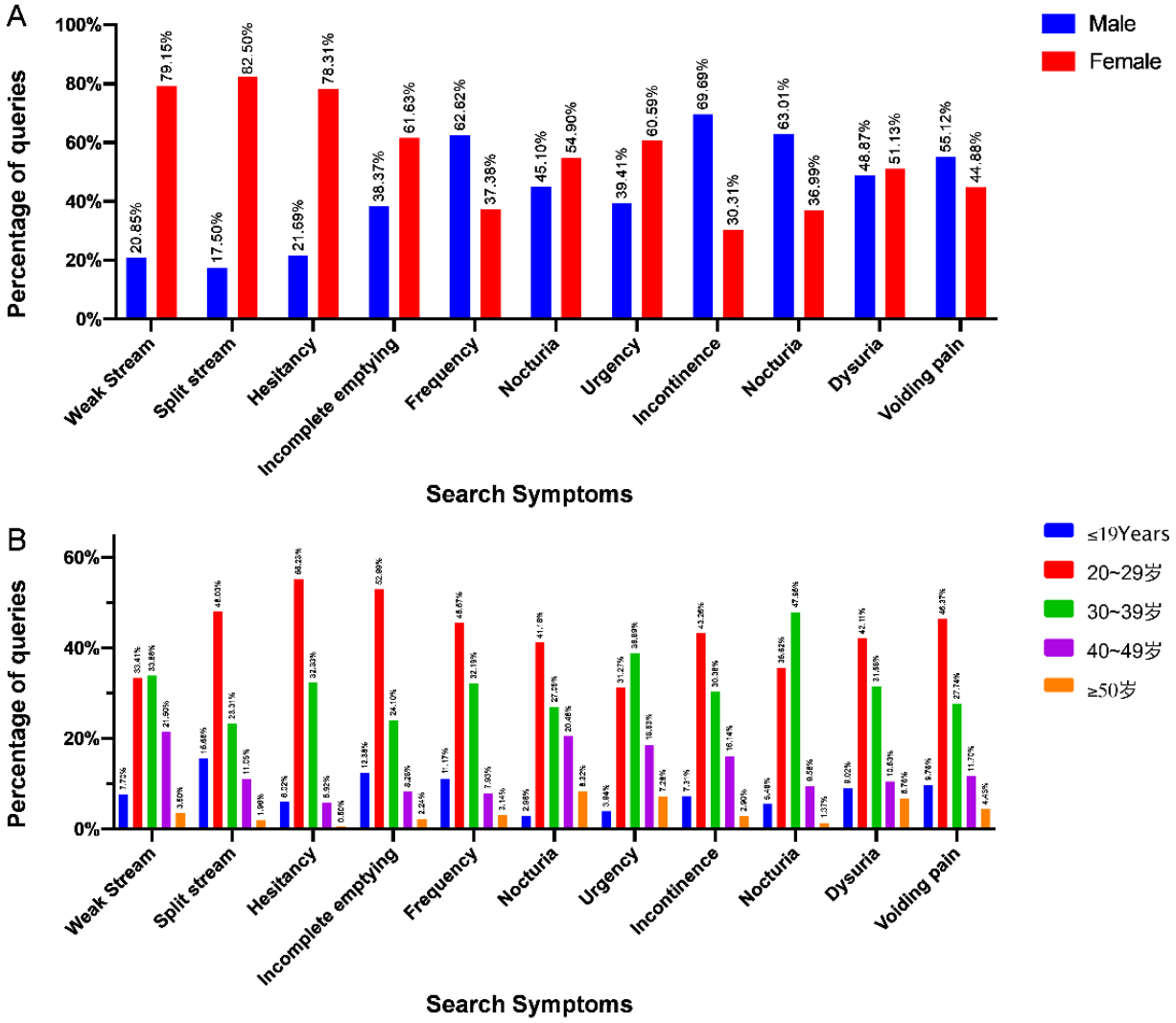


Demographic Differences

The age distribution of the population for all LUTS-related inquiries showed that individuals aged 20 to 40 years made 73.86% (49,218,123/66,635,247) of the total search inquiries. Further, individuals aged 40 to 49 years made 12.29% (8,193,922/66,635,247) of the total search inquiries for all LUTS-related terms. People aged over 50 years accounted for the least amount of general search inquiries. With regard to gender differences, searches for the term *urinary frequency*, *incontinence*, *nocturia*, and *voiding pain* were more prevalent in male populations. Other terms, such as *weak stream*, *split stream*, and *hesitancy*, were mainly inquired by female populations (Figure 4).

XSL•FO
RenderX

Figure 4. Demographic distributions of lower urinary tract symptoms–related searches. A: Gender distribution. B: Age distribution.



Search Incentives

LUTS are chronic conditions that are common in adult men and are frequently associated with BPH, chronic prostatitis, cystitis, urethritis, and urinary tract infection. Hence, we examined the correlation among the daily BSIs of these terms in order to explore people's motivation for conducting LUTS-related searches. It was revealed that most of the searches for LUTS-related terms were related to those for urinary diseases to varying degrees (Table 1). There was no significant correlation between BPH and *weak stream* (P<.11) and between chronic prostatitis and *dysuria* (P=.053).

XSL•FO
RenderX

Table 1. Correlations between lower urinary tract symptoms–related search terms and related diagnoses (N=3543).

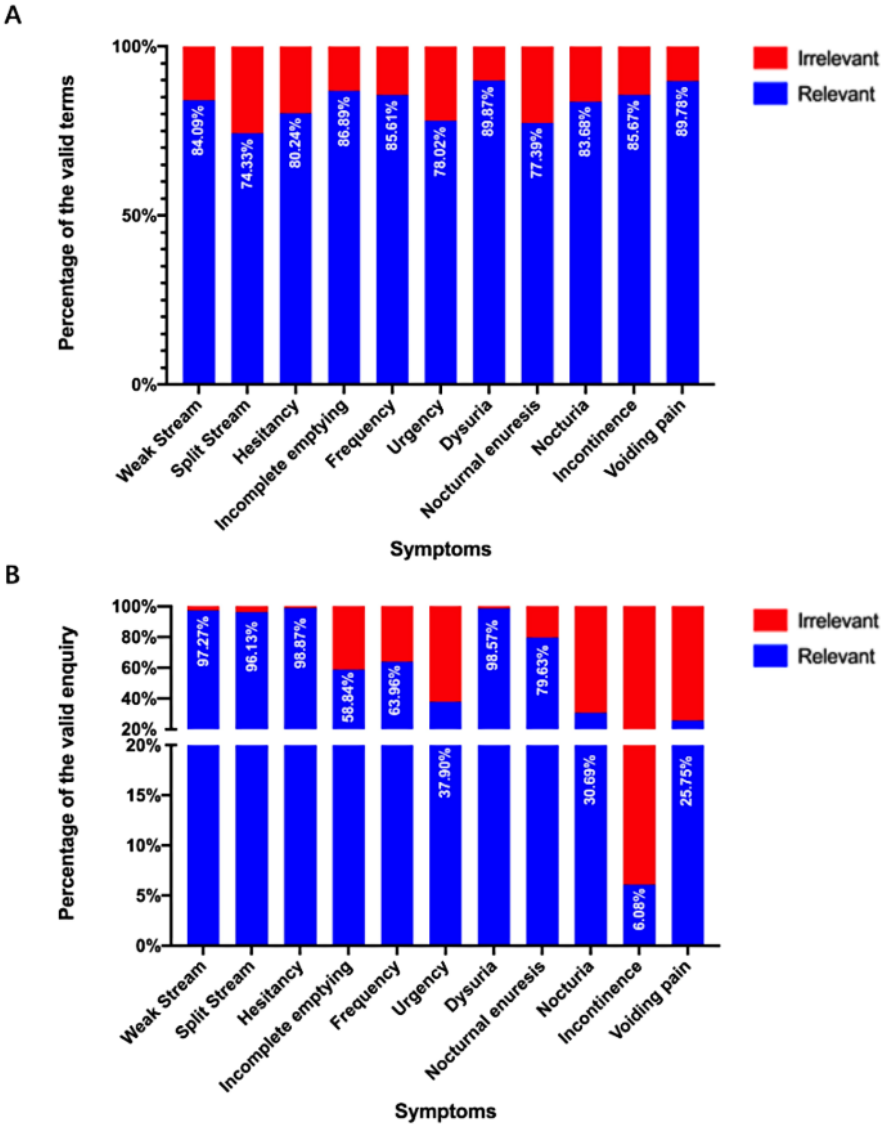
Search terms	Benign prostatic hyperplasia		Chronic prostatitis		Cystitis		Urethritis		Urinary tract infection	
	r	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value	*r*	*P* value
Weak stream	0.011	.11	−0.240	<.001	0.344	<.001	0.098	<.001	−0.106	<.001
Split stream	0.527	<.001	0.230	<.001	0.313	<.001	0.490	<.001	0.664	<.001
Hesitancy	−0.218	<.001	−0.219	<.001	0.324	<.001	−0.145	<.001	−0.287	<.001
Incomplete emptying	0.636	<.001	0.378	<.001	0.064	<.001	0.443	<.001	0.608	<.001
Nocturia	0.657	<.001	0.407	<.001	0.139	<.001	0.325	<.001	0.659	<.001
Dysuria	0.222	<.01	−0.033	.053	0.197	<.001	0.255	<.001	0.159	<.001
Incontinence	0.707	<.001	0.053	<.001	0.061	<.001	0.531	<.001	0.677	<.001
Urinary frequency	0.624	<.001	0.492	<.001	0.086	<.001	0.386	<.001	0.598	<.001
Nocturnal enuresis	0.498	<.001	0.305	<.001	0.330	<.001	0.196	<.001	0.519	<.001
Urgency	0.071	<.001	0.083	<.001	0.273	<.001	0.085	<.001	−0.188	<.001
Odynuria	−0.305	<.001	−0.291	<.001	−0.359	<.001	−0.166	<.001	−0.279	<.001

Relative Terms of Keywords and Search Frequency

We reviewed and sorted the top searched relative terms of keywords in the Baidu Index platform. In order to classify the relative terms based on the main concerns of the users, 13 hypothetical (used to describe users' concerns more explicitly and comprehensively) categories of domains and their core meanings were defined. These categories were as follows: irrelevant, symptoms, ethology quests, treatment, medical information, diagnoses, products and hospitals, diagnosis confirmation, tests and examinations, prognosis, traditional Chinese medicine–conceptualized quests related to diagnoses, symptoms, and treatment materials. Detailed percentages of the relevant search terms and search frequencies are listed in Figure 5. In Figure 6 and Figure 7, the inquiry terms and frequencies related to LUTS in Baidu Index are categorized and presented with their percentages. Furthermore, the top 3 terms for each domain are listed in Multimedia Appendix 6, along with their BSIs.

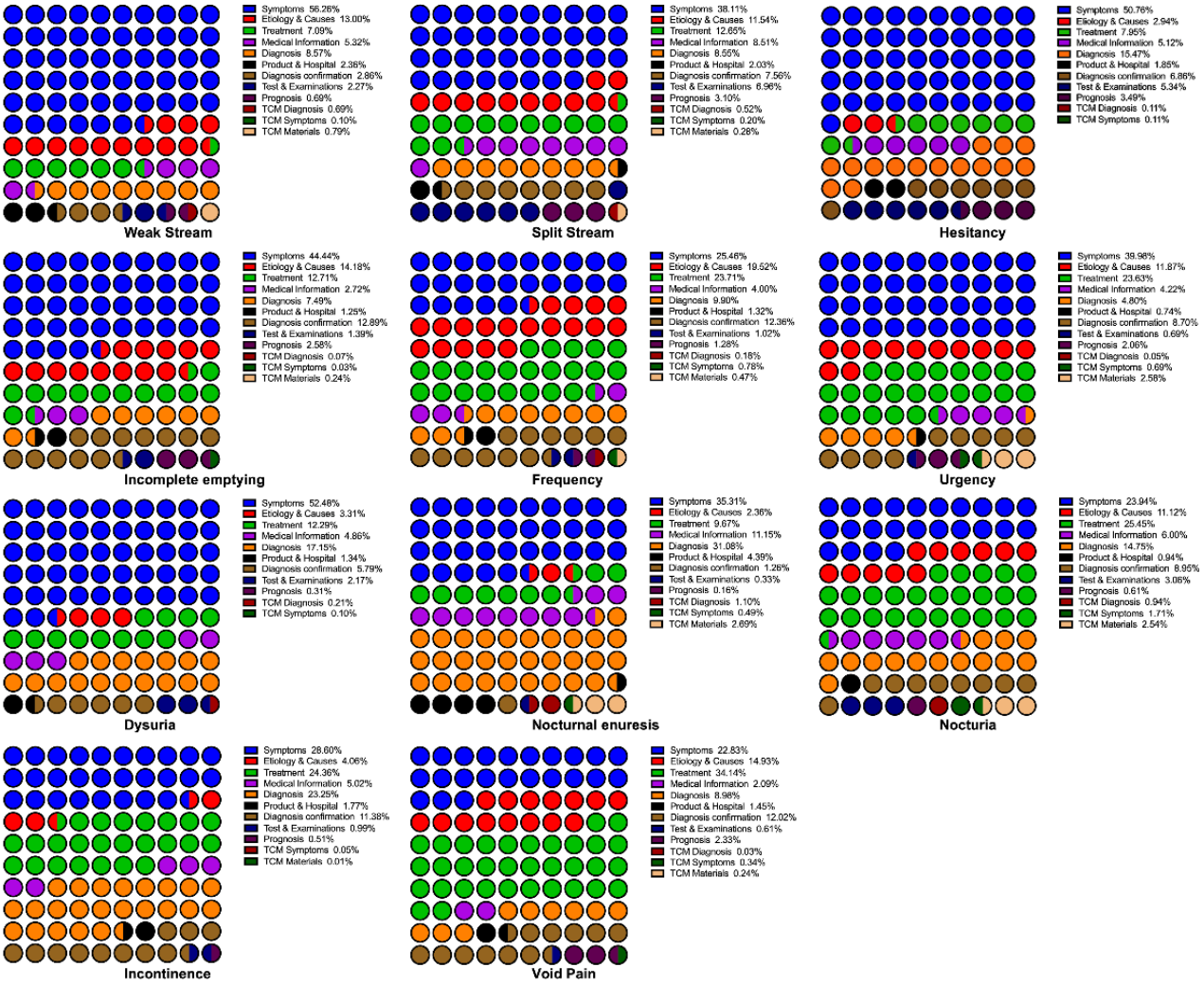
XSL•FO
RenderX

Figure 5. The percentage of valid searches for lower urinary tract symptoms–related terms. A: The percentage of valid term inquiries. B: The frequency of valid term inquiries.



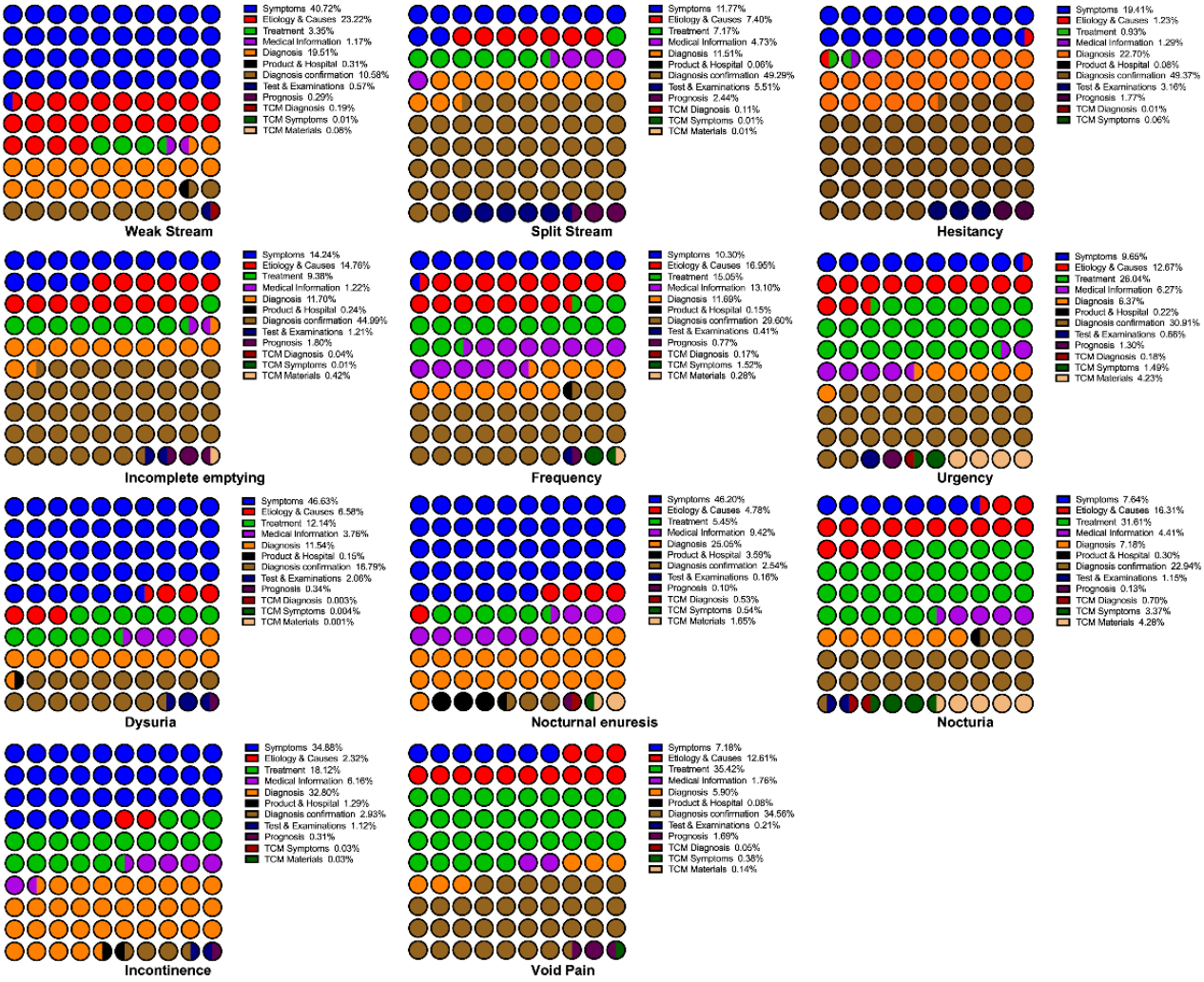
XSL•FO
RenderX

Figure 6. The percentage of term inquiry categories in Baidu Index related to lower urinary tract symptoms. TCM: traditional Chinese medicine.



XSL•FO
RenderX

Figure 7. Percentage of enquiry frequency categories related to LUTS in Baidu Index. LUTS: Lower Urinary Tract Symptoms, TCM: Traditional Chinese Medicine.



Discussion

Principal Findings

Our findings revealed that internet search trends in Baidu Index could comprehensively reflect the actual concerns of Chinese populations regarding health issues. Previous studies have confirmed that users' information-seeking behavior profiles from big data platforms are adeptly proficient in forecasting disease outbreaks, identifying interest at the population level, monitoring health education campaigns, and tracking the trends of patients' preferences [14,15,24,25]. As a primary health care information source, patients are more inclined to initiate a preliminary consult on the internet before seeing a doctor [34]. Hence, investigations of the infodemiology and infoveillance features of LUTS-related terms are capable of revealing a massive amount of users' behavioral information and reflecting actual prevalence trends to some extent [24,25]. This feature is particularly crucial when there are a lack of related epidemiology data and real-time information.

By using Baidu Index, we found that the popularity of each entry varied greatly. The most inquired term, *urinary frequency*, was searched 25 times more often than the least inquired term, *nocturnal enuresis*. Intuitively, this result shows different degrees of attention for LUTS and therefore reflects the relative prevalence of symptoms in the population. With regard to trends, we illustrated that search volumes for each LUTS-related term had maintained a certain order of magnitude. A sustained growing trend was observed for the terms *frequency*, *incomplete emptying*, *nocturia*, and *incontinence*. The trends for other phrases were relatively flat and steady. These trends indicate that while the prevalence of LUTS may remain stable, public awareness and internet penetration may be on the rise, and people's habits of seeking medical advice are changing. Further, 75.31% (50,183,464/66,635,247) of the search requests were conducted on mobile devices. This fact may have resulted from the development of portable technology and mobile internet. Some clinics use smartphones to track older adults with frailty and their quality of life [35]. Additionally, 43.5% of individuals have acknowledged that they use cellphones while on the toilet; therefore, it is no trouble for these users to conduct searches while "on the go," especially when they are experiencing a LUTS attack [36].

We noticed a spike in search volume for the term *incontinence*, which surged to over 3400 from 1100 on May 6, 2013. However, just 15 days prior, an earthquake (7.0 on the Richter scale) struck

XSL•FO
RenderX

Ya'an, Sichuan Province, resulting in 196 deaths, 21 missing individuals, and 11,470 injured individuals. We believe that the spike in search volume for *incontinence* is related to this event, as most incontinence symptoms would have appeared after the injuries from the earthquake [37]. Additionally, we noticed an abnormal fluctuation that was above the average search volume for the term *split stream*. The total increase in this term's search volume originated from term 4 (尿尿分叉; ie, *Wee Wee split*). We suspect that this may have been caused by the bidding manipulation of specific keywords by some individuals or institutions because the popularity of other terms has not increased significantly [25]. However, due to the openness of the Baidu Index platform, this abnormal fluctuation can only affect the search popularity of a specific term for a certain period of time [23]. Nevertheless, it will not affect the popularity of other synonyms. Additionally, despite the fact that the competitive ranking promotion mechanism is well known in Baidu's product line, this mechanism is limited to ranking search results instead of guiding or changing users' search preferences by shifting their needs. Hence, this abnormal search volume phenomenon only occurs occasionally. The relatively consistent search trend for other terms suggests that the BSI can still be a good indicator of public attention toward specific LUTS if keywords are sufficiently included in the search.

After analyzing the regional differences in search volumes, we found that the search trends of each LUTS-related term were in line with those of other infodemiology research [24,25]. The search volume for each LUTS-related term was highest in the East China region, followed by those in North, South, and Central China. In Northwest China, search volumes were lowest. However, the trends revealed by the BSI contradict previous investigations on the prevalence of LUTS and BPH. In earlier investigations, the prevalence of LUTS was highest in the northwest and northeast regions and lowest in the southwest areas [12,38]. However, at the provincial level, the prevalence of LUTS in Guangdong, Shanghai, and Beijing was ranked highest compared to those of other provinces [38]. These three provinces are representative of the South, East, and North China regions. Further, the data sets of previous investigations pooled individuals aged over 40 years. As a result of this, the rankings from these investigations were not much different from ours. Moreover, medical infrastructures, internet access, and public health awareness are considered better in these higher income regions. Coupled with the fact that Baidu's data sources involve people of all ages from across the whole country, the geographic patterns in BSIs may reflect the socioeconomic and population rankings of regions in mainland China.

In Groutz et al's [39] and Heylen et al's [40] investigations, the prevalence of voiding difficulty in the female population was vastly underestimated. The prevalence of voiding difficulty significantly increases with age and the degree of pelvic organ prolapse [39,40]. Additionally, voiding symptoms such as weak streams, hesitancy, and strained urination are especially prevalent in females when other comorbidities are present [41]. Specifically, the pooled incidence rate of voiding difficulty is 21.3% and 15.5% in males and females, respectively [42]. Thus symptom differences based on sex are not significant. Further, when comparing LUTS prevalence rates, Apostolidis et al [43] recognized that they failed to investigate incidents of incontinence in both sexes. This was because certain symptoms, such as incontinence, are not listed in the IPSS or National Institutes of Health Chronic Prostatitis Symptom Index assessments, which are the most commonly applied assessments and checklists for men with LUTS and chronic pelvic pain syndrome symptoms [42]. The IPSS is universally used, even when evaluating the efficacy of prostate surgery—a known leading cause of male urinary incontinence [44,45]. The evaluations for male incontinence is listed in the "expanded" index for "prostate cancer patients" [46]. This is probably because of the belief that male incontinence has almost always resulted from prolonged bladder outlet obstruction [47]. Nevertheless, it should be noted that other causes, such as sphincter injury following prostatic surgery, polypharmacy, detrusor degeneration, and sustained urinary tract infection, have also been identified as common factors of male incontinence onset [48]. Therefore, prostate problems may be an essential factor of the systematic neglect of primary incontinence symptoms in the male population. Similarly, the diagnostic criteria for female bladder obstruction have been misestimated, and it is recommended that female bladder obstruction should be confirmed with urodynamic examinations [39,43]. Despite these issues, there are differences in the distribution of LUTS between men and women. It is worth pointing out that the real factor that determines whether patients seek medical treatment is not the characteristics of symptoms but the severity of symptoms [49]. Additionally, as internet users with web-based health issue inquiries are 60% more likely to seek and use health care services, the ratio of BSIs for reported symptoms could be used to remind practitioners to avoid ignoring "uncommon" symptoms in each gender population [50].

Baidu Index lists the top 10 related theme words for each search term. In this platform, the trends for correlated search terms are calculated and updated weekly. This system enables practitioners to gain insight into the most concerning problems from patients, confirm individuals' main intentions, and examine the most exposed information to users. In our study, we found that the related theme words were not limited to the field of disease diagnosis, cause, and treatment, as previously described [25]. Instead, users had an extensive range of questions and concerns about LUTS (eg, the best hospital, treatment efficacies, and disease prognosis), and Baidu Index includes a lot of content that has nothing to do with health counseling. The differences in the distribution of the theme words varied vastly for each LUTS-related term. In terms of the relevance ratio, the term *dysuria* had the highest number of relevant term entries and search frequencies. However, although the term *urinary incontinence* was included in 85.67% (8134/9495) of the number of valid search entries, this term had the lowest search volume (1,726,303/283,932,352, 6.08%). Further, we noticed that the terms related to symptoms, etiology, and treatment were the terms that were the most inquired by Baidu platform users. With regard to the term *incontinence*, a large number of inquiries related to diagnosis mainly included the term *incontinence*, which is both a symptom description and a diagnostic. Hence it is clear that Baidu platform users are more inclined to make inquiries by describing their symptoms and seek treatment.

XSL•FO
RenderX

Although the diagnosis and prognosis of LUTS are also of concern, the rate is relatively low.

We also noticed that there were numerous inquiries that were formatted as “what are the symptoms of...?” This form of question implies that Baidu platform users may have a definite awareness of the diagnoses related to their symptoms and that they have needed to confirm or exclude a particular diagnosis. This is a typical internet self-assessment pattern. Further, for the treatment category, we found that most users focused on inquiries such as “the most effective medication” or “the fastest efficacy.” These facts reveal that Baidu platform users’ views of LUTS and related diagnoses are one-sided. Their intention is to control LUTS as soon as possible instead of consulting a doctor to obtain a definite diagnosis and treatment regimen. It is worth pointing out that despite the fact that diseases indicated by LUTS are often not life-threatening, completing routine urine tests or ultrasonic examinations as part of standardized diagnosis and treatment procedures can at least reduce the likelihood of misdiagnosing malignancies [51,52]. Furthermore, some patients believe that examinations with negative results are conducted with the purpose of making a profit and are therefore meaningless and a waste of money [53]. These beliefs may worsen when these patients come across information that claims people can cure LUTS with free, “do-it-yourself” methods. Such information will strengthen their prejudice against hospitals and increase their likeliness of using “do-it-yourself” methods to heal their LUTS [54]. Patients’ limited understanding of health issues predisposes them to misdiagnoses, results in illness delays, and poses considerable health risks.

Limitations

Several limitations in this study should be addressed. One is that Baidu Index only analyzes search data from Baidu and does not analyze search data from social media platforms. Therefore, the data generated by these platforms could not be assessed. Further, the types of Baidu platform users could not be determined due to confidentiality (ie, users’ privacy protection). Consequently, the analysis of demographic data regarding information-seeking preferences and behaviors could only be based on age, gender, and regions. Information such as socioeconomic status, ethnicity, and educational background were not obtainable. Furthermore, LUTS and their associated diseases, such as prostatitis, BPH, and cystitis, are not covered by national surveillance systems and lack corresponding epidemiological data. Additionally, instead of real search frequencies, the BSI is just a weighted index derivative; we could only speculate the relative popularity of LUTS-related search terms and their development trends.

Notwithstanding these limitations, this is the first study that investigates public concerns about LUTS, to the best of our knowledge. We chose to examine the infodemiology characteristics of LUTS because the clinical manifestations of LUTS are obvious to patients. As such, LUTS are easy to compare when analyzing clinical diagnoses and diseases. Therefore, research on the search volumes of LUTS-related keywords can directly mirror existing problems from the patient perspective.

Finally, we believe that infodemiology research based on the BSI reveals people's behaviors when they search for related keywords and that such data can be used as an important reference for understanding the population’s needs. In terms of medical and health-related information, Baidu should seek government health management departments' help to revise and standardize medical-related keywords and consultations, so that their platform can better serve consumers, potential patients, health management departments, and medical practitioners. Since the real-time data on search volumes and information-seeking behaviors on the platform are renewed over time, these data can potentially be used to provide supplementary references and up-to-date information for improving care practice standards and making policies in a timelier manner.

Conclusion

Web-based interest in LUTS-related terms fluctuated wildly and was reflected timely by Baidu Index in mainland China. Significant variability was observed in the web-based search popularity of each LUTS-related term, and popularity differed based on personal interests, the population’s concerns, regional variations, and gender differences. These data can be used by medical professionals to track the prevalence of LUTS and the population’s interests, guide the establishment of disease-specific health care policies, and optimize physician-patient health care sessions.

Authors' Contributions

SW and LJ developed the protocol and project; collected, managed, and analyzed the data; and wrote and edited the manuscript. MM, CW, XW, and BY collected and managed the data and wrote and edited the manuscript. MS developed the protocol and project, analyzed the data, and wrote and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of symptom domains and term confirmation. BSI: Baidu search index; ICS: International Continence Society; LUTS: lower urinary tract symptoms.
[PNG File , 19 KB - jmir_v23i7e27029_app1.png]

XSL•FO
RenderX

Multimedia Appendix 2
List of keywords used in the composite search index.
[PDF File (Adobe PDF File), 269 KB - jmir_v23i7e27029_app2.pdf]

Multimedia Appendix 3
Web-based interest in lower urinary tract symptoms domains over the last 10 years.
[PDF File (Adobe PDF File), 691 KB - jmir_v23i7e27029_app3.pdf]

Multimedia Appendix 4
Joinpoint graphic of Baidu search index trends for each each lower urinary tract symptoms domain.
[PNG File , 777 KB - jmir_v23i7e27029_app4.png]

Multimedia Appendix 5
Regional distribution of web-based interest for each lower urinary tract symptoms domain over the last 10 years. A: Regional rates for each area. B: Annual Baidu search index trends for each region.
[PDF File (Adobe PDF File), 518 KB - jmir_v23i7e27029_app5.pdf]

Multimedia Appendix 6
List of the most inquired terms for each lower urinary tract symptoms domain.
[PDF File (Adobe PDF File), 1027 KB - jmir_v23i7e27029_app6.pdf]

References

1. Abrams P, Cardozo L, Fall M, Griffiths D, Rosier P, Ulmsten U, Standardisation Sub-committee of the International Continence Society. The standardisation of terminology of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. Neurourol Urodyn 2002;21(2):167-178. [doi: 10.1002/nau.10052] [Medline: 11857671]
2. Coyne KS, Sexton CC, Kopp ZS, Ebel-Bitoun C, Milsom I, Chapple C. The impact of overactive bladder on mental health, work productivity and health-related quality of life in the UK and Sweden: results from EpiLUTS. BJU Int 2011 Nov;108(9):1459-1471. [doi: 10.1111/j.1464-410X.2010.10013.x] [Medline: 21371240]
3. Speakman M, Kirby R, Doyle S, Ioannou C. Burden of male lower urinary tract symptoms (LUTS) suggestive of benign prostatic hyperplasia (BPH) - focus on the UK. BJU Int 2015 Apr;115(4):508-519. [doi: 10.1111/bju.12745] [Medline: 24656222]
4. Jhaveri J, Gauthier-Loiselle M, Gagnon-Sanschagrin P, Wu EQ. The economic burden of nocturia on the U.S. health care system and society: A national health and nutrition examination survey analysis. J Manag Care Spec Pharm 2019 Dec;25(12):1398-1408. [doi: 10.18553/jmcp.2019.19191] [Medline: 31566054]
5. Coyne KS, Sexton CC, Thompson CL, Milsom I, Irwin D, Kopp ZS, et al. The prevalence of lower urinary tract symptoms (LUTS) in the USA, the UK and Sweden: results from the Epidemiology of LUTS (EpiLUTS) study. BJU Int 2009 Aug;104(3):352-360. [doi: 10.1111/j.1464-410X.2009.08427.x] [Medline: 19281467]
6. Lukhnova LI, Tleugabylova AM, Khodzhaeva LU, Okladnikova IL. [The bactericidal activity of aqueous solutions of pervomur]. Lab Delo 1990(11):77-78. [Medline: 1704971]
7. Coyne KS, Margolis MK, Kopp ZS, Kaplan SA. Racial differences in the prevalence of overactive bladder in the United States from the epidemiology of LUTS (EpiLUTS) study. Urology 2012 Jan;79(1):95-101. [doi: 10.1016/j.urology.2011.09.010] [Medline: 22055692]
8. Mourad S, Shokeir A, Ayoub N, Ibrahim M, Reynolds N, Donde S, et al. Prevalence and impact of lower urinary tract symptoms: Results of the epic survey in Egypt. Neurourol Urodyn 2019 Feb;38(2):637-643. [doi: 10.1002/nau.23875] [Medline: 30575129]
9. Griebling TL. Worldwide prevalence estimates of lower urinary tract symptoms, overactive bladder, urinary incontinence, and bladder outlet obstruction. BJU Int 2011 Oct;108(7):1138-1139. [doi: 10.1111/j.1464-410X.2011.10498.x] [Medline: 21914110]
10. Wang Y, Hu H, Xu K, Wang X, Na Y, Kang X. Prevalence, risk factors and the bother of lower urinary tract symptoms in China: a population-based survey. Int Urogynecol J 2015 Jun;26(6):911-919. [doi: 10.1007/s00192-015-2626-8] [Medline: 25653032]
11. Wang J, Liao L, Liu M, Sumarsono B, Cong M. Epidemiology of lower urinary tract symptoms in a cross-sectional, population-based study: The status in China. Medicine (Baltimore) 2018 Aug;97(34):e11554 [FREE Full text] [doi: 10.1097/MD.0000000000011554] [Medline: 30142751]
12. Zhang W, Zhang X, Li H, Wu F, Wang H, Zhao M, et al. Prevalence of lower urinary tract symptoms suggestive of benign prostatic hyperplasia (LUTS/BPH) in China: results from the China Health and Retirement Longitudinal Study. BMJ Open 2019 Jun 19;9(6):e022792 [FREE Full text] [doi: 10.1136/bmjopen-2018-022792] [Medline: 31221864]

XSL•FO
RenderX

13. Zhang L, Zhu L, Xu T, Lang J, Li Z, Gong J, et al. A population-based survey of the prevalence, potential risk factors, and symptom-specific bother of lower urinary tract symptoms in adult Chinese women. Eur Urol 2015 Jul;68(1):97-112. [doi: 10.1016/j.eururo.2014.12.012] [Medline: 25572826]
14. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
15. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
16. Bousquet J, Agache I, Anto JM, Bergmann KC, Bachert C, Annesi-Maesano I, et al. Google Trends terms reporting rhinitis and related topics differ in European countries. Allergy 2017 Aug;72(8):1261-1266. [doi: 10.1111/all.13137] [Medline: 28140507]
17. Brusselaers N, Monstrey S, Vogelaers D, Hoste E, Blot S. Severe burn injury in Europe: a systematic review of the incidence, etiology, morbidity, and mortality. Crit Care 2010;14(5):R188 [FREE Full text] [doi: 10.1186/cc9300] [Medline: 20958968]
18. Hartnett KP, Kite-Powell A, Patel MT, Haag BL, Sheppard MJ, Dias TP, et al. Syndromic surveillance for e-cigarette, or vaping, product use-associated lung injury. N Engl J Med 2020 Feb 20;382(8):766-772. [doi: 10.1056/NEJMsr1915313] [Medline: 31860794]
19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
20. Helft M, Barboza D. Google shuts China site in dispute over censorship. The New York Times. 2010 Mar 23. URL: https://www.nytimes.com/2010/03/23/technology/23google.html [accessed 2021-06-21]
21. Danny F. Paradoxes, Google and China - How censorship can harm and intellectual property can harness innovation. In: Lopez-Tarruella A, editor. Google and the Law. The Hague, The Netherlands: T. M. C. Asser Press; Feb 22, 2012.
22. Worldwide desktop market share of leading search engines from January 2010 to February 2021. Statista. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2021-06-21]
23. Zhao C, Yang Y, Wu S, Wu W, Xue H, An K, et al. Search trends and prediction of human brucellosis using Baidu index data from 2011 to 2018 in China. Sci Rep 2020 Apr 03;10(1):5896 [FREE Full text] [doi: 10.1038/s41598-020-62517-7] [Medline: 32246053]
24. Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, et al. Association between cancer incidence and mortality in web-based data in China: Infodemiology study. J Med Internet Res 2019 Jan 29;21(1):e10677 [FREE Full text] [doi: 10.2196/10677] [Medline: 30694203]
25. Wang T, Xia Q, Chen X, Jin X. Use of Baidu Index to track Chinese online behavior and interest in kidney stones. Risk Manag Healthc Policy 2020 Jul 03;13:705-712. [doi: 10.2147/RMHP.S245822] [Medline: 32694922]
26. The 46th China Statistical Report on Internet Development. China Internet Network Information Center (CNNIC), China. 2020. URL: https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/ [accessed 2021-06-15]
27. 中国网民科普需求搜索行为报告(2019年第一季度). China Association for Science and Technology. 2019 Apr 26. URL: https://www.cast.org.cn/art/2019/4/26/art_1281_94546.html [accessed 2021-06-15]
28. D'Ancona C, Haylen B, Oelke M, Abranches-Monteiro L, Arnold E, Goldman H, et al. The International Continence Society (ICS) report on the terminology for adult male lower urinary tract and pelvic floor symptoms and dysfunction. Neurourol Urodyn 2019 Feb;38(2):433-477. [doi: 10.1002/nau.23897] [Medline: 30681183]
29. Gajewski JB, Schurch B, Hamid R, Averbeck M, Sakakibara R, Agrò EF, et al. An International Continence Society (ICS) report on the terminology for adult neurogenic lower urinary tract dysfunction (ANLUTD). Neurourol Urodyn 2018 Mar;37(3):1152-1161. [doi: 10.1002/nau.23397] [Medline: 29149505]
30. Hashim H, Blanker MH, Drake MJ, Djurhuus JC, Meijlink J, Morris V, et al. International Continence Society (ICS) report on the terminology for nocturia and nocturnal lower urinary tract function. Neurourol Urodyn 2019 Feb;38(2):499-508. [doi: 10.1002/nau.23917] [Medline: 30644584]
31. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. PLoS Negl Trop Dis 2017 Mar 06;11(3):e0005354. [doi: 10.1371/journal.pntd.0005354] [Medline: 28263988]
32. The definition of seasons in China. China Meteorological Administration. 2013. URL: http://www.cma.gov.cn/2011xzt/2013zhuant/20130620_3/2013062002/201308/t20130816_223401.html [accessed 2021-06-21]
33. Mayo-Yáñez M, Calvo-Henríquez C, Chiesa-Estomba C, Lechien JR, González-Torres L. Google Trends application for the study of information search behaviour on oropharyngeal cancer in Spain. Eur Arch Otorhinolaryngol 2021 Jul;278(7):2569-2575. [doi: 10.1007/s00405-020-06494-7] [Medline: 33237476]
34. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]

XSL•FO
RenderX

35. Fontecha J, Navarro FJ, Hervás R, Bravo J. Elderly frailty detection by using accelerometer-enabled smartphones and clinical information records. Pers Ubiquitous Comput 2012 May 12;17(6):1073-1083 [FREE Full text] [doi: 10.1007/s00779-012-0559-5]
36. Banawas S, Abdel-Hadi A, Alaidarous M, Alshehri B, Bin Dukhyil AA, Alsaweed M, et al. Multidrug-resistant bacteria associated with cell phones of healthcare professionals in selected hospitals in Saudi Arabia. Can J Infect Dis Med Microbiol 2018 Dec 24;2018:6598918 [FREE Full text] [doi: 10.1155/2018/6598918] [Medline: 30675320]
37. Zhou TJ. [A renal function survey in paraplegic patients of the Tangshan earthquake]. Zhonghua Wai Ke Za Zhi 1989 Aug;27(8):482-483, 510. [Medline: 2620605]
38. Wang W, Guo Y, Zhang D, Tian Y, Zhang X. The prevalence of benign prostatic hyperplasia in mainland China: evidence from epidemiological surveys. Sci Rep 2015 Aug 26;5:13546 [FREE Full text] [doi: 10.1038/srep13546] [Medline: 26306721]
39. Groutz A, Blaivas JG, Chaikin DC. Bladder outlet obstruction in women: definition and characteristics. Neurourol Urodyn 2000;19(3):213-220. [doi: 10.1002/(sici)1520-6777(2000)19:3<213::aid-nau2>3.0.co;2-u] [Medline: 10797578]
40. Haylen BT, Krishnan S, Schulz S, Verity L, Law M, Zhou J, et al. Has the true prevalence of voiding difficulty in urogynecology patients been underestimated? Int Urogynecol J Pelvic Floor Dysfunct 2007 Jan;18(1):53-56. [doi: 10.1007/s00192-006-0094-x] [Medline: 16596458]
41. Lee SH, Lee JG, Min GE, Lee H, Lee CH, Yoo KH. Usefulness of total bladder capacity and post-void residual urine volume as a predictor of detrusor overactivity with impaired contractility in stroke patients. Exp Ther Med 2012 Dec;4(6):1112-1116 [FREE Full text] [doi: 10.3892/etm.2012.708] [Medline: 23226784]
42. Akikwala TV, Fleischman N, Nitti VW. Comparison of diagnostic criteria for female bladder outlet obstruction. J Urol 2006 Nov;176(5):2093-2097. [doi: 10.1016/j.juro.2006.07.031] [Medline: 17070266]
43. Apostolidis A, Kirana P, Chiu G, Link C, Tsiouprou M, Hatzichristou D. Gender and age differences in the perception of bother and health care seeking for lower urinary tract symptoms: results from the hospitalised and outpatients' profile and expectations study. Eur Urol 2009 Dec;56(6):937-947. [doi: 10.1016/j.eururo.2009.07.050] [Medline: 19683857]
44. Huang S, Tsai C, Tseng C, Shih M, Yeh Y, Chien K, et al. Comparative efficacy and safety of new surgical treatments for benign prostatic hyperplasia: systematic review and network meta-analysis. BMJ 2019 Nov 14;367:l5919 [FREE Full text] [doi: 10.1136/bmj.l5919] [Medline: 31727627]
45. Simone G, Misuraca L, Anceschi U, Minisola F, Ferriero M, Guaglianone S, et al. Urethra and ejaculation preserving robot-assisted simple prostatectomy: Near-infrared fluorescence imaging-guided Madigan technique. Eur Urol 2019 Mar;75(3):492-497. [doi: 10.1016/j.eururo.2018.11.051] [Medline: 30553615]
46. Chang P, Szymanski KM, Dunn RL, Chipman JJ, Litwin MS, Nguyen PL, et al. Expanded prostate cancer index composite for clinical practice: development and validation of a practical health related quality of life instrument for use in the routine clinical care of patients with prostate cancer. J Urol 2011 Sep;186(3):865-872 [FREE Full text] [doi: 10.1016/j.juro.2011.04.085] [Medline: 21788038]
47. Hampson LA, Elliott SP, Erickson BA, Vanni AJ, Myers JB, McClung C, TURNS. Multicenter analysis of urinary urgency and urge incontinence in patients with anterior urethral stricture disease before and after urethroplasty. J Urol 2016 Dec;196(6):1700-1705 [FREE Full text] [doi: 10.1016/j.juro.2016.08.015] [Medline: 27521692]
48. Hester AG, Kretschmer A, Badlani G. Male incontinence: The etiology or basis of treatment. Eur Urol Focus 2017 Oct;3(4-5):377-384. [doi: 10.1016/j.euf.2017.12.002] [Medline: 29249687]
49. Ho L, Chu PS, Consigliere DT, Zainuddin ZM, Bolong D, Chan C, et al. Symptom prevalence, bother, and treatment satisfaction in men with lower urinary tract symptoms in Southeast Asia: a multinational, cross-sectional survey. World J Urol 2018 Jan;36(1):79-86. [doi: 10.1007/s00345-017-2097-2] [Medline: 29051978]
50. Ybarra ML, Suman M. Help seeking behavior and the internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
51. Engeler DS, Baranowski AP, Dinis-Oliveira P, Elneil S, Hughes J, Messelink EJ, European Association of Urology. The 2013 EAU guidelines on chronic pelvic pain: is management of chronic pelvic pain a habit, a philosophy, or a science? 10 years of development. Eur Urol 2013 Sep;64(3):431-439. [doi: 10.1016/j.eururo.2013.04.035] [Medline: 23684447]
52. Naber KG, Bonkat G, Wagenlehner FME. The EAU and AUA/CUA/SUFU guidelines on recurrent urinary tract infections: What is the difference? Eur Urol 2020 Nov;78(5):645-646. [doi: 10.1016/j.eururo.2020.06.032] [Medline: 32616407]
53. Young B, Fogarty AW, Skelly R, Shaw D, Thurley P, Sturrock N, et al. Patients' attitudes towards cost feedback to doctors to prevent unnecessary testing: a qualitative focus group study. Public Health 2020 Aug;185:338-340 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.023] [Medline: 32726730]
54. Lavorgna L, De Stefano M, Sparaco M, Moccia M, Abbadessa G, Montella P, et al. Fake news, influencers and health-related professional participation on the Web: A pilot study on a social-network of people with Multiple Sclerosis. Mult Scler Relat Disord 2018 Oct;25:175-178 [FREE Full text] [doi: 10.1016/j.msard.2018.07.046] [Medline: 30096683]

Abbreviations

APC: annual percent change
BPH: benign prostatic hyperplasia

XSL•FO
RenderX

BSI: Baidu search index
ICS: International Continence Society
IPSS: International Prostate Symptom Score
LUTS: lower urinary tract symptoms

Edited by R Kukafka; submitted 08.01.21; peer-reviewed by H Ma, X Hu, P Wang, A Alasmari; comments to author 11.02.21; revised version received 10.03.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Wei S, Ma M, Wu C, Yu B, Jiang L, Wen X, Fu F, Shi M
Using Search Trends to Analyze Web-Based Interest in Lower Urinary Tract Symptoms-Related Inquiries, Diagnoses, and Treatments in Mainland China: Infodemiology Study of Baidu Index Data
J Med Internet Res 2021;23(7):e27029
URL: https://www.jmir.org/2021/7/e27029
doi:10.2196/27029
PMID:34255683

©Shanzun Wei, Ming Ma, Changjing Wu, Botao Yu, Lisha Jiang, Xi Wen, Fudong Fu, Ming Shi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Prediction of Asthma Hospitalizations for the Common Cold Using Google Trends: Infodemiology Study

Bernardo Sousa-Pinto[1,2], MD, PhD; Jaana I Halonen[3], PhD; Aram Antó[4], MSc; Vesa Jormanainen[3], MD; Wienczyslawa Czarlewski[4,5,6], PhD; Anna Bedbrook[4,6], BSc; Nikolaos G Papadopoulos[7,8], PhD; Alberto Freitas[1,2], PhD; Tari Haahtela[9], MD; Josep M Antó[10,11,12,13], PhD; João Almeida Fonseca[1,2], MD, PhD; Jean Bousquet[6,14,15,16], PhD

[1]MEDCIDS – Department of Community Medicine, Information and Health Decision Sciences; Faculty of Medicine, University of Porto, Porto, Portugal

[2]CINTESIS – Center for Health Technology and Services Research; University of Porto, Porto, Portugal

[3]Finnish Institute for Health and Welfare (THL), Helsinki, Finland

[4]MASK-air, Montpellier, France

[5]Medical Consulting Czarlewski, Levallois, France

[6]MACVIA-France, Montpellier, France

[7]Allergy Department, 2nd Pediatric Clinic, Athens General Children's Hospital "P&A Kyriakou", University of Athens, Athens, Greece

[8]Division of Infection, Immunity & Respiratory Medicine, Royal Manchester Children's Hospital, University of Manchester, Manchester, United Kingdom

[9]Skin and Allergy Hospital, Helsinki University Hospital, and University of Helsinki, Helsinki, Finland

[10]ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain

[11]Universitat Pompeu Fabra, Barcelona, Spain

[12]IMIM - Hospital del Mar Medical Research Institute, Barcelona, Spain

[13]CIBER Epidemiología y Salud Pública - CIBERESP, Barcelona, Spain

[14]Charité, Universitätsmedizin Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

[15]Comprehensive Allergy Center, Department of Dermatology and Allergy, Berlin Institute of Health, Berlin, Germany

[16]University Hospital, Montpellier, France

Corresponding Author:
Bernardo Sousa-Pinto, MD, PhD
MEDCIDS – Department of Community Medicine, Information and Health Decision Sciences; Faculty of Medicine, University of Porto
Rua Dr. Plácido da Costa s/n
Porto
Portugal
Phone: 351 225513600
Email: bernardosousapinto@protonmail.com

Abstract

Background: In contrast to air pollution and pollen exposure, data on the occurrence of the common cold are difficult to incorporate in models predicting asthma hospitalizations.

Objective: This study aims to assess whether web-based searches on *common cold* would correlate with and help to predict asthma hospitalizations.

Methods: We analyzed all hospitalizations with a main diagnosis of asthma occurring in 5 different countries (Portugal, Spain, Finland, Norway, and Brazil) for a period of approximately 5 years (January 1, 2012-December 17, 2016). Data on web-based searches on *common cold* were retrieved from Google Trends (GT) using the *pseudo-influenza syndrome* topic and local language search terms for *common cold* for the same countries and periods. We applied time series analysis methods to estimate the correlation between GT and hospitalization data. In addition, we built autoregressive models to forecast the weekly number of asthma hospitalizations for a period of 1 year (June 2015-June 2016) based on admissions and GT data from the 3 previous years.

Results: In time series analyses, GT data on *common cold* displayed strong correlations with asthma hospitalizations occurring in Portugal (correlation coefficients ranging from 0.63 to 0.73), Spain (ρ=0.82-0.84), and Brazil (ρ=0.77-0.83) and moderate correlations with those occurring in Norway (ρ=0.32-0.35) and Finland (ρ=0.44-0.47). Similar patterns were observed in the

XSL•FO
RenderX

correlation between forecasted and observed asthma hospitalizations from June 2015 to June 2016, with the number of forecasted hospitalizations differing on average between 12% (Spain) and 33% (Norway) from observed hospitalizations.

Conclusions: Common cold–related web-based searches display moderate-to-strong correlations with asthma hospitalizations and may be useful in forecasting them.

(J Med Internet Res 2021;23(7):e27044) doi:10.2196/27044

KEYWORDS

asthma; common cold; Google Trends; hospitalizations; time series analysis; mobile phone

Introduction

Background

Asthma poses a substantial burden on health care, with hospitalizations being one of the main drivers of asthma-related costs [1]. The prediction of asthma hospitalization patterns may take into account major risk factors for asthma exacerbations, such as occurrence of the common cold (most often due to rhinovirus infections) [2-4], air pollution, and pollen exposure. However, although air pollution and pollen peaks can be measured (allowing for alert systems to be developed), data on rhinovirus infections are more difficult to obtain and thus to be incorporated into prediction models.

Infodemiology data open new possibilities for the development of models predicting asthma hospitalizations. Infodemiology is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [5,6]. Infodemiology comprises *supply-based* and *demand-based* approaches, with the latter including the analysis of web-based searches to assess individuals' health-seeking behavior [6]. Google Trends (GT) is one of the most frequently used tools to assess trends in web-based searches. This Google service displays the relative volume of searches for which a keyword (or set of keywords) is entered into the Google search engine [7]. Web-based searches on *asthma* and related terms have been assessed in previous studies [8,9]. However, GT for the search term *asthma* [7] only allows for the easy identification of large outbreaks (such as thunderstorm-induced asthma) [10] or media coverage–driven search peaks [11]. In fact, seasonal variations in pollen concentrations, which influence the occurrence of asthma exacerbations, are not reflected in GT for *asthma* [9]. However, to date, no study has examined the relationship between GT for *common cold*, the major risk factor for asthma exacerbations, and asthma hospitalizations.

Objectives

In this study, we aim to assess whether asthma hospitalizations could be predicted by GT along with data from previous hospital admissions. To do so, we aim (1) to assess and discuss GT for *common cold* in different countries of the world, (2) to correlate GT for *common cold* and hospitalization data, and (3) to build models forecasting asthma hospitalizations for a period of 1 year (based on GT and past hospitalization data), correlating observed and predicted asthma hospitalizations.

Methods

Study Design

We conducted an infodemiology study to (1) correlate GT on rhinovirus-related search terms with asthma hospitalizations for a period of approximately 5 years (2012-2016) in Portugal, Spain, Finland, Norway, and Brazil and (2) assess whether such GT, along with previous admissions, were able to predict asthma hospitalizations. This study complies with the methodological framework of Mavragani and Ochoa [12].

Queries and Data Sources for GT for Rhinovirus-Related Search Terms

GT topics are groups of search terms that concern the same concept, irrespective of their language [9,12-14]. This is particularly relevant, as there are countries with several words and idioms referring to the *common cold* (a paradigmatic example is Spain, a country with five official and co-official languages) as well as others with ambiguous words referring to the *common cold* (eg, English-speaking countries, in which *cold* has a double meaning). Assessing data from several countries, we found that *common cold* was listed as a *pseudo-influenza syndrome* topic (this topic was renamed as *common cold* at the time of manuscript submission).

We then retrieved country-level GT on rhinovirus-related search terms from January 1, 2012, to December 17, 2016, in Portugal, Spain, Finland, Norway, and Brazil. The countries were selected according to the possibility of having nationwide weekly asthma hospitalization data available for comparison. To provide a wider perspective, we plotted GT patterns for the *pseudo-influenza syndrome* topic, not only in these 5 countries, in which asthma hospitalizations were also assessed, but also in 11 additional countries.

The assessed time frame was selected considering both GT and asthma hospitalization data. On the one hand, for periods longer than 5 years, GT data are presented on a monthly level rather than on a weekly level (with monthly intervals being insufficiently sensitive for assessing variations in asthma hospitalizations). On the other hand, a complete period of 5 years was not assessed (with the last 2 weeks of December 2016 not being assessed) on account of available asthma hospitalization data (see the subsection *Asthma Hospitalization Data Sources*).

For each country, we tested two different GT queries: (1) one consisting of the *pseudo-influenza syndrome* topic, and (2) another being a combination of search terms consisting of words for *common cold* (selected on discussion with native speakers

XSL•FO
RenderX

of each language). The search term combinations were as follows:

- Portugal: *constipação* + *resfriado*
- Spain: *resfriado* + *resfrío* + *catarro* + *constipado* + *refredat* + *constipate* + *arrefriado* + *hotzeri*
- Finland: *flunssa* + *nuha* + *vilustuminen*
- Norway: *forkjølelse* + *forkjøling* + *snue* + *krimsjuke*
- Brazil: *resfriado*

Quotation marks were not used because each keyword consisted of a single word. Misspellings or nonaccentuated forms were not included in the search term combinations. In Portugal, the words without diacritical marks (ie, *constipacao* or *constipaçao*) have a much lower relative volume of searches (with many *zero-value* observations) than the correct word *constipação*, with identical relative search volumes being observed whenever misspelled words are or are not included in search term combinations. In Spain, identical relative search volumes are also observed, whether or not the misspelled word *resfrio* is included in search term combinations. In Norway, the misspelled variants *forkjolelse* and *forkjoling* generate negligible results (ie, all observations with a relative search volume of zero).

State-level analyses were also performed in Spain and Brazil. For Spain, we performed separate analyses for the three most populous autonomous communities (Andalusia, Catalonia, and Madrid). For Brazil, we separately analyzed data from the most populous state in three of the five Brazilian geographical regions (as defined by the Brazilian Institute of Geography and Statistics): São Paulo (Southeast), Rio Grande do Sul (South), and Bahia (Northeast). For the two other geographical regions (North and Central-West), the most populous states had low-quality GT data (with many weeks recorded as 0).

GT was accessed via its web interface. Categories and subcategories were not selected in our searches. GT data sources, other than *web searches*, have not been used. Searches and data extraction were performed on January 13, 2020, with a single data extraction for each country.

Asthma Hospitalization Data Sources

In the 5 studied countries, we assessed all hospitalizations with asthma as the main diagnosis (ie, International Classification of Diseases, Ninth Revision, Clinical Modification code 493.x or International Classification of Diseases, Tenth Revision, code J45) occurring in public hospitals from January 1, 2012, to December 17, 2016 (the last 2 weeks of 2016 were not included, as we did not have any information on 2017 Portuguese and Brazilian discharges, and many patients admitted at the end of 2016 were discharged in 2017). Hospitalization data were retrieved from (1) the Hospital Morbidity database (provided by the Portuguese Central Administration of the Healthcare System) for Portugal, (2) the Hospital Morbidity Survey databases (*Encuesta de morbilidad hospitalaria, Instituto Nacional de Estadistica*) for Spain, (3) the National Hospital Discharge Register (*Hoitoilmoitusrekisteri*, HILMO) for Finland, (4) the Norwegian Patient Registry (*Norsk Pasientregister*) for Norway, and (5) DATASUS data from the Single Health System (*Sistema Único de Saúde*) for Brazil.

Statistical Analysis

In brief, we performed two major types of analyses. First, we assessed the correlations between GT data and asthma hospitalizations in each country after applying time series analysis methods. Subsequently, we built models forecasting asthma hospitalizations for a period of 1 year based on GT and hospitalization data from the previous 3 years. To test for the predictive ability of the models, forecasted and observed asthma admissions were compared. Both GT and hospitalization data were presented weekly. The performance of analysis on a weekly basis allowed for the detection of short-term variations without the large random fluctuations that can be observed when data are analyzed on a daily basis.

In detail, we calculated Pearson correlation coefficients to assess the correlation between GT data and asthma hospitalizations in each country. In addition, we performed cross-correlation analysis because (1) for GT data, a relevant trend is expected, mirroring an increase in Google searches with the passing of years, and (2) GT results are expressed as relative search values (ie, percentages in relation to the maximum observed value of the whole period), whereas hospitalizations are expressed as absolute values. The time series can be decomposed into three components: trend, seasonal effects, and random errors. We removed the trend component for both GT and hospitalization data and then estimated cross-correlation coefficients between GT and hospitalization data of the same week and with different week lags (namely, with 1, 2, 3, and 4 weeks of difference between GT data and hospitalization data, to assess whether search volumes displayed better correlation with asthma hospitalizations occurring in subsequent weeks than with those occurring in the same week).

We built seasonal autoregressive integrated moving average (ARIMA) models to forecast variations in asthma hospitalizations over a period of 1 year. We started in the Northern Hemisphere Summer of 2015 (week of June 21, 2015) and based our models on the trend component of asthma hospitalizations and on GT data from the 3 previous years (weeks of July 1, 2012-June 14, 2015). Seasonal ARIMA models are defined by the parameters $(p, d, q)(P, D, Q)_s$, where p corresponds to the order of autoregression, d is the degree of difference, q is the order of the moving average part, P is the seasonal order of autoregression, D is the seasonal integration, Q is the seasonal moving average, and s is the length of the seasonal period [11,15]. We applied seasonal ARIMA $(3,0,2)(0,1,1)_{52}$ models. Such models were chosen, among others with p=3 or p=4 (as suggested by the time series autocorrelation function plots) and with other parameters, on account of their lower corrected Akaike information criteria and nondetection of correlated residuals (both as assessed by the Ljung-Box test and by the autocorrelation function plots, with no significant spikes being observed). Asthma hospitalizations and GT data from July 2012 to June 2015 were used as a training set, with hospitalizations between the weeks of June 21, 2015, and June 19, 2016 being forecasted based on observed GT for that period. The predictive ability of the models was assessed by calculating (1) the correlation coefficients between the predicted variation in hospitalizations and the observed trend in hospitalizations

XSL•FO
RenderX

for that period, (2) the correlation coefficients between the predicted variation in hospitalizations and the actual number of asthma hospitalizations (ie, without time series decomposition) for that period, (3) the average weekly difference between the numbers of predicted and observed hospitalizations, and (4) the number of weeks whose number of observed asthma hospitalizations fell outside the 95% CI for predicted admissions.

Normality was assessed from the skewness and kurtosis of each distribution (with values lower than −1 or higher than 1 indicating deviation from normality). The 95% CIs of the correlation coefficients were computed to assess their precision and determine whether they were significantly different from 0. All analyses were performed using the R software, version 4.0.0 (R Foundation for Statistical Computing).

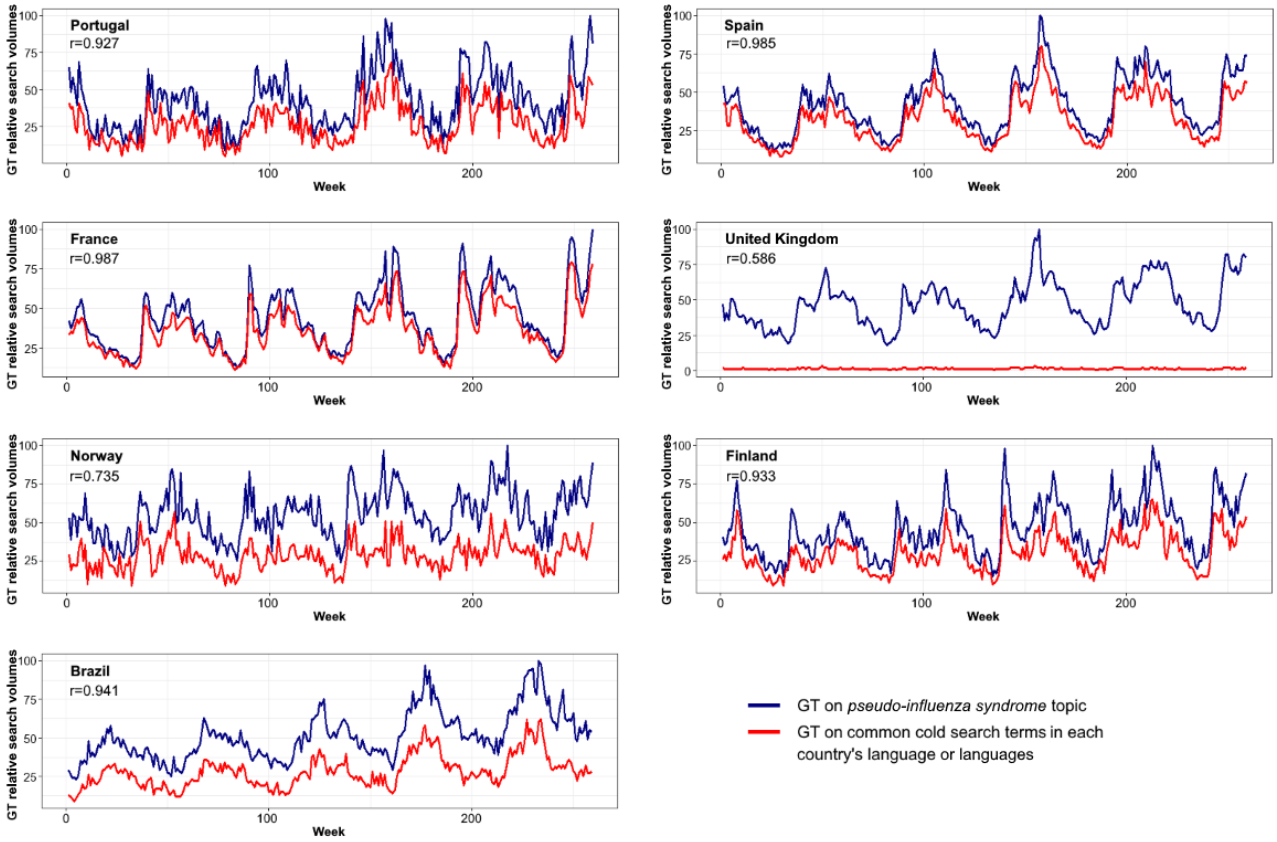
Results

Overview

Assessing data from several countries, we observed that *common cold* was listed as *pseudo-influenza syndrome* topic, although with variable correlation between GT on the *pseudo-influenza syndrome* topic and *common cold* words (Figure 1). A relatively low correlation coefficient was observed for the United Kingdom, as we compared the *pseudo-influenza syndrome* topic with the search expression *common cold*, which may not be frequently used, particularly when compared with the term *cold*. However, we did not query the latter term because of its double meaning.

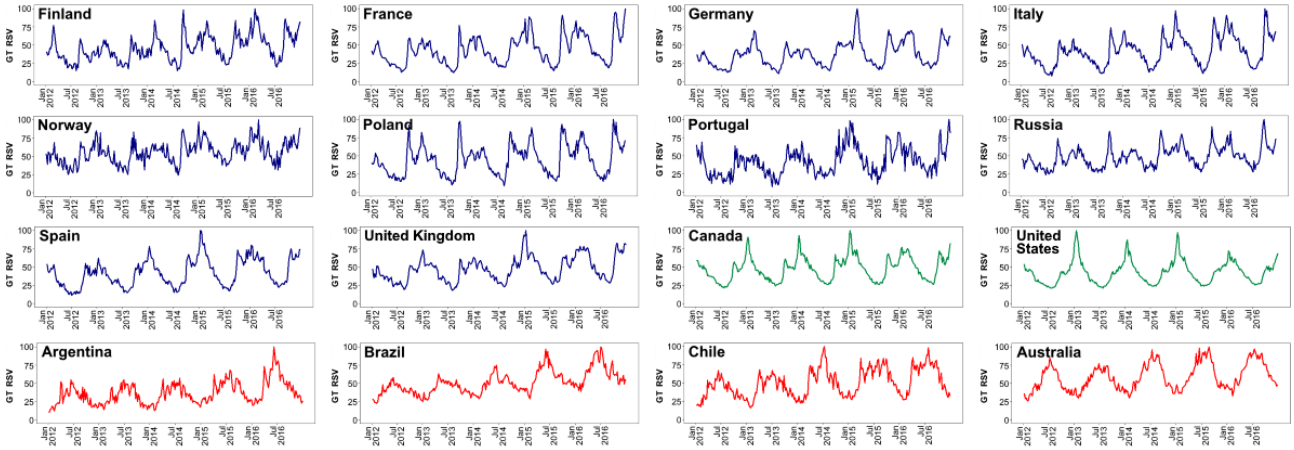
Between 2012 and 2016, GT data for *pseudo-influenza syndrome* presented similar patterns across 16 countries for which GT data were plotted, with peaks in the winter and valleys in the summer (Figure 2). In the 5 main assessed countries, correlations between untransformed GT on the *pseudo-influenza syndrome* topic and asthma hospitalizations varied between 0.10 (in the Brazilian state of Bahia) and 0.69 (in Spain; Table 1). Similar values were observed when analyzing the correlations between GT and words for the common cold.

Figure 1. Google Trends data on the *pseudo-influenza syndrome* topic and on *common cold* search terms in each country's respective language or languages (r: Pearson correlation coefficient). For the United Kingdom, r=0.769 when GT data on *pseudo-influenza syndrome* and on *common cold* search terms are retrieved separately. GT: Google Trends.



XSL•FO
RenderX

Figure 2. Google Trends data on *pseudo-influenza syndrome* for 16 countries in Europe (blue), North America (green), and the Southern Hemisphere (red) for a period of 5 years (2012-2016). GT: Google Trends; RSV: relative search volume.



XSL•FO
RenderX

Table 1. Correlation and cross-correlation coefficients between common cold Google Trends data (ie, Google Trends data on the pseudo-influenza syndrome topic and on common cold search terms) and asthma hospitalizations for the period 2012-2016.

Country or region	Correlation coefficients (95% CI) based on observed data	Cross-correlation coefficients (95% CI) after removal of the trend component				
		Week lag[a]-0	Week lag[a]-1	Week lag[a]-2	Week lag[a]-3	Week lag[a]-4
***Pseudo-influenza syndrome* topic**						
Portugal	0.54 (0.42 to 0.66)	0.68 (0.54 to 0.81)	0.73 (0.59 to 0.86)	0.67 (0.53 to 0.81)	0.71 (0.57 to 0.85)	0.65 (0.52 to 0.79)
Spain	0.69 (0.57 to 0.81)	0.83 (0.69 to 0.97)	0.84 (0.71 to 0.98)	0.83 (0.69 to 0.96)	0.80 (0.67 to 0.94)	0.76 (0.62 to 0.89)
Andalusia	0.54 (0.42 to 0.66)	0.63 (0.50 to 0.77)	0.68 (0.55 to 0.82)	0.68 (0.54 to 0.81)	0.67 (0.54 to 0.81)	0.66 (0.52 to 0.80)
Catalonia	0.65 (0.53 to 0.78)	0.80 (0.66 to 0.93)	0.80 (0.66 to 0.93)	0.79 (0.65 to 0.93)	0.79 (0.65 to 0.92)	0.74 (0.60 to 0.87)
Madrid	0.63 (0.51 to 0.75)	0.67 (0.53 to 0.80)	0.69 (0.55 to 0.82)	0.65 (0.52 to 0.79)	0.65 (0.52 to 0.79)	0.62 (0.48 to 0.76)
Finland	0.16 (0.04 to 0.29)	0.44 (0.30 to 0.58)	0.44 (0.30 to 0.57)	0.42 (0.29 to 0.56)	0.32 (0.18 to 0.46)	0.25 (0.11 to 0.39)
Norway	0.15 (0.03 to 0.27)	0.32 (0.18 to 0.45)	0.35 (0.21 to 0.49)	0.26 (0.12 to 0.39)	0.18 (0.05 to 0.32)	0.18 (0.05 to 0.32)
Brazil	0.26 (0.14 to 0.39)	0.83 (0.69 to 0.97)	0.77 (0.64 to 0.91)	0.70 (0.57 to 0.84)	0.62 (0.49 to 0.76)	0.54 (0.40 to 0.67)
São Paulo	0.45 (0.33 to 0.57)	0.66 (0.52 to 0.80)	0.58 (0.45 to 0.72)	0.44 (0.30 to 0.58)	0.33 (0.19 to 0.47)	0.28 (0.14 to 0.42)
Rio Grande do Sul	0.54 (0.42 to 0.66)	0.67 (0.53 to 0.80)	0.64 (0.50 to 0.77)	0.59 (0.45 to 0.72)	0.57 (0.44 to 0.71)	0.53 (0.40 to 0.67)
Bahia	0.10 (−0.02 to 0.22)	0.50 (0.37 to 0.64)	0.49 (0.36 to 0.63)	0.54 (0.40 to 0.68)	0.47 (0.33 to 0.61)	0.40 (0.26 to 0.53)
***Common cold* search terms**						
Portugal	0.53 (0.41 to 0.65)	0.63 (0.50 to 0.77)	0.68 (0.54 to 0.82)	0.68 (0.54 to 0.81)	0.65 (0.51 to 0.79)	0.59 (0.45 to 0.72)
Spain	0.69 (0.57 to 0.82)	0.82 (0.69 to 0.96)	0.84 (0.70 to 0.97)	0.82 (0.68 to 0.96)	0.80 (0.66 to 0.94)	0.75 (0.62 to 0.89)
Andalusia	0.55 (0.43 to 0.67)	0.62 (0.48 to 0.76)	0.65 (0.51 to 0.78)	0.66 (0.53 to 0.80)	0.65 (0.52 to 0.79)	0.66 (0.52 to 0.79)
Catalonia	0.67 (0.55 to 0.79)	0.78 (0.65 to 0.92)	0.78 (0.65 to 0.92)	0.78 (0.64 to 0.92)	0.76 (0.62 to 0.90)	0.71 (0.58 to 0.85)
Madrid	0.61 (0.49 to 0.73)	0.63 (0.50 to 0.77)	0.66 (0.52 to 0.80)	0.64 (0.50 to 0.78)	0.64 (0.50 to 0.78)	0.62 (0.48 to 0.76)
Finland	0.24 (0.12 to 0.36)	0.47 (0.34 to 0.61)	0.46 (0.32 to 0.59)	0.40 (0.26 to 0.54)	0.32 (0.19 to 0.46)	0.25 (0.11 to 0.38)
Norway	0.22 (0.10 to 0.35)	0.35 (0.21 to 0.49)	0.33 (0.20 to 0.47)	0.24 (0.11 to 0.38)	0.15 (0.02 to 0.29)	0.15 (0.02 to 0.29)
Brazil	0.37 (0.25 to 0.49)	0.82 (0.69 to 0.96)	0.77 (0.63 to 0.91)	0.69 (0.55 to 0.82)	0.61 (0.47 to 0.74)	0.52 (0.38 to 0.65)
São Paulo	0.46 (0.34 to 0.58)	0.67 (0.54 to 0.81)	0.60 (0.47 to 0.74)	0.46 (0.33 to 0.60)	0.37 (0.24 to 0.51)	0.30 (0.16 to 0.43)
Rio Grande do Sul	0.55 (0.43 to 0.67)	0.61 (0.48 to 0.75)	0.57 (0.43 to 0.70)	0.53 (0.40 to 0.67)	0.52 (0.38 to 0.65)	0.46 (0.33 to 0.60)
Bahia	0.18 (0.06 to 0.30)	0.40 (0.26 to 0.54)	0.43 (0.30 to 0.57)	0.40 (0.26 to 0.54)	0.35 (0.21 to 0.48)	0.29 (0.15 to 0.43)

[a]Week lag corresponds to the week difference between Google Trends and hospitalization data (eg, a week lag of 1 implies that Google Trends data of a certain week will be correlated with hospitalization data of the following week).

Time Series Results

In time series analyses, GT on the *pseudo-influenza syndrome* topic correlated more strongly with asthma hospitalizations occurring in the subsequent week than with those occurring in the same week in Portugal (ρ=0.73 vs ρ=0.68), Spain (ρ=0.84 vs ρ=0.83), and Norway (ρ=0.35 vs ρ=0.32) but not in Finland (ρ=0.44 in both cases) or Brazil (ρ=0.77 vs ρ=0.83; Table 1; Figure 3). Similar results were observed with GT on *common cold* search terms (Portugal: ρ=0.68 vs ρ=0.63; Spain: ρ=0.84 vs ρ=0.82; Finland: ρ=0.47 vs ρ=0.46; Norway: ρ=0.32 vs ρ=0.35; Brazil: ρ=0.77 vs ρ=0.82). Relevant regional differences were observed in Spain and Brazil. In Spain, stronger correlations were observed in Catalonia than in Madrid or Andalusia. In Brazil, stronger correlations were observed in Rio Grande do Sul and São Paulo than in Bahia.

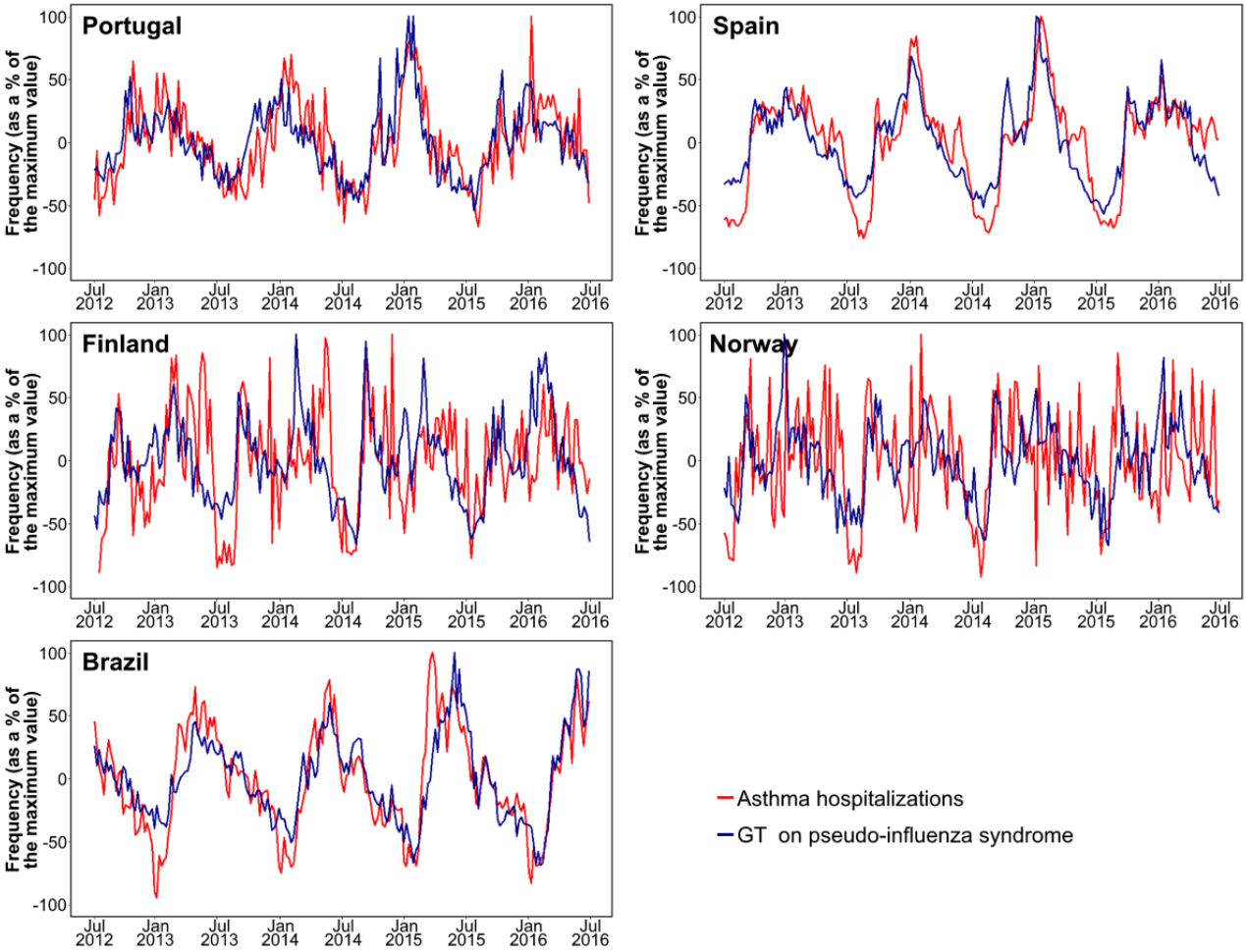
Forecasts for 1-year (June 2015 to June 2016) variations in asthma hospitalizations obtained through seasonal ARIMA models strongly correlated with actual observed asthma hospitalizations for the same period in Spain (ρ=0.88-0.91), Brazil (ρ=0.87-0.94), and Portugal (ρ=0.69-0.79; Table 2). Such correlations were moderate for Finland (ρ=0.49-0.55) and Norway (ρ=0.37-0.45). In Spain, the strongest correlations were observed for Catalonia (ρ=0.86-0.87), whereas in Brazil, they were observed for Rio Grande do Sul (ρ=0.89-0.91).

From June 2015 to June 2016, we also forecasted the number of asthma hospitalizations occurring each week and compared it with the number of observed asthma hospitalizations (Table 3; Figures 4 and 5). The weekly number of predicted hospitalizations showed, on average, a 12% difference compared with the number of observed asthma hospitalizations in Spain.

XSL•FO
RenderX

This difference was 23% in Portugal, 16%-17% in Finland, 32%-33% in Norway, and 21%-23% in Brazil. In 1 year, the number of weeks in which the absolute frequency of observed asthma hospitalizations did not fall within the predicted 95% CI ranged between 0 (Rio Grande do Sul) and 8 (Brazil as a whole).

Figure 3. Google Trends data on *pseudo-influenza syndrome* and asthma hospitalizations (2012-2016) in Portugal, Spain, Finland, Norway, and Brazil. The trend component of time series has been plotted after removal of the seasonal effects and random error components. GT: Google Trends.



XSL•FO
RenderX

Table 2. Results of forecasts for 1-year variations in asthma hospitalizations: correlation coefficients between predicted variations in asthma hospitalizations and actually observed asthma hospitalizations over 1 year (June 2015 to June 2016).

Country or region	Transformed observed hospitalizations[a], correlation coefficient (95% CI)	Untransformed observed hospitalizations[b], correlation coefficient (95% CI)
***Pseudo-influenza syndrome* topic**		
Portugal	0.79 (0.67-0.88)	0.74 (0.60-0.84)
Spain	0.90 (0.82-0.95)	0.88 (0.79-0.93)
Andalusia	0.75 (0.61-0.85)	0.75 (0.60-0.85)
Catalonia	0.87 (0.77-0.93)	0.86 (0.77-0.92)
Madrid	0.83 (0.72-0.90)	0.82 (0.70-0.89)
Finland	0.54 (0.26-0.73)	0.49 (0.19-0.71)
Norway	0.37 (0.09-0.60)	0.41 (0.15-0.64)
Brazil	0.93 (0.89-0.96)	0.87 (0.80-0.92)
São Paulo	0.68 (0.45-0.84)	0.67 (0.51-0.79)
Rio Grande do Sul	0.89 (0.83-0.94)	0.91 (0.86-0.95)
Bahia	0.85 (0.74-0.92)	0.81 (0.72-0.89)
***Common cold* search terms**		
Portugal	0.76 (0.63-0.85)	0.69 (0.58-0.79)
Spain	0.91 (0.85-0.95)	0.88 (0.77-0.95)
Andalusia	0.79 (0.67-0.87)	0.78 (0.64-0.88)
Catalonia	0.87 (0.81-0.93)	0.86 (0.79-0.92)
Madrid	0.84 (0.74-0.91)	0.83 (0.72-0.90)
Finland	0.55 (0.26-0.75)	0.49 (0.18-0.72)
Norway	0.39 (0.16-0.58)	0.45 (0.19-0.63)
Brazil	0.94 (0.90-0.96)	0.88 (0.82-0.92)
São Paulo	0.73 (0.50-0.88)	0.72 (0.59-0.81)
Rio Grande do Sul	0.89 (0.82-0.94)	0.90 (0.84-0.95)
Bahia	0.85 (0.77-0.92)	0.81 (0.71-0.89)

[a]Correlation coefficients between predicted weekly asthma hospitalization trends and actual observed hospitalizations after applying time series analysis methods (ie, after removing the trend component).

[b]Correlation coefficients between predicted weekly hospitalization trends and actual observed raw numbers of weekly asthma hospitalizations.

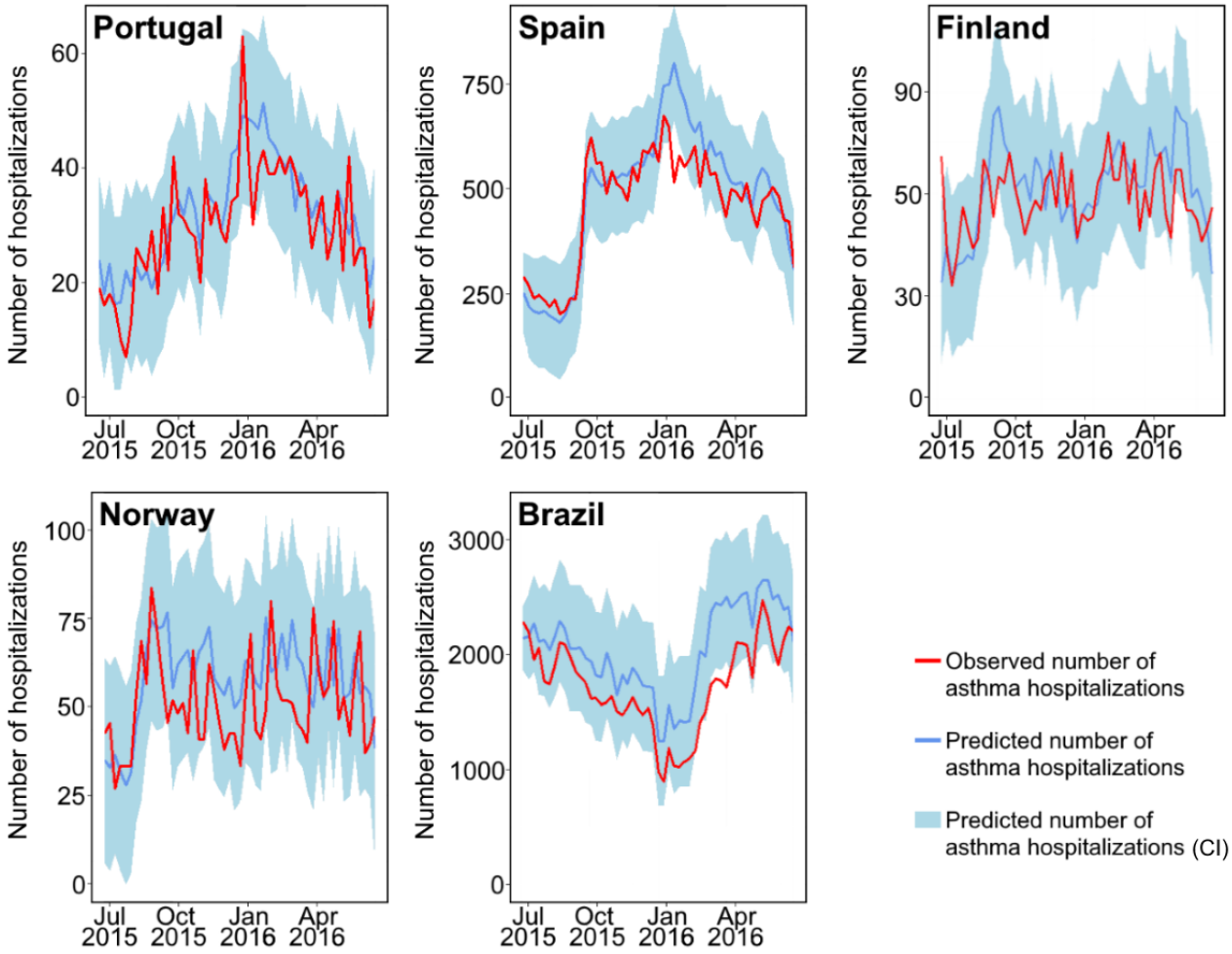
XSL•FO
RenderX

Table 3. Results of 1-year (June 2015-June 2016) forecasts for the number of asthma hospitalizations based on autoregressive integrated moving average models including common cold–related Google Trends data and asthma hospitalizations of the previous 3 years.

Country or region	Correlation (95% CIs) between number of predicted and observed hospitalizations	Average difference in the absolute numbers of predicted and observed weekly hospitalizations, N (average % difference)	Weeks with observed hospitalizations outside predicted 95% CIs, n (%)
***Pseudo-influenza syndrome* topic**			
Portugal	0.79 (0.68-0.87)	5 (23.3)	1 (1.9)
Spain	0.92 (0.85-0.96)	54 (11.6)	4 (7.5)
Andalusia	0.76 (0.62-0.85)	9 (21.5)	5 (9.4)
Catalonia	0.88 (0.80-0.92)	13 (16.3)	5 (9.4)
Madrid	0.80 (0.67-0.88)	15 (21.1)	6 (11.3)
Finland	0.45 (0.16-0.69)	10 (16.7)	4 (7.5)
Norway	0.40 (0.17-0.59)	14 (31.8)	1 (1.9)
Brazil	0.88 (0.80-0.92)	328 (20.6)	7 (13.2)
São Paulo	0.63 (0.40-0.80)	52 (32.3)	3 (5.7)
Rio Grande do Sul	0.88 (0.82-0.93)	22 (22.8)	1 (1.9)
Bahia	0.79 (0.67-0.87)	63 (24.4)	5 (9.4)
***Common cold* search terms**			
Portugal	0.77 (0.65-0.86)	6 (22.5)	1 (1.9)
Spain	0.90 (0.85-0.96)	54 (11.6)	4 (7.5)
Andalusia	0.78 (0.66-0.88)	9 (23.5)	6 (11.3)
Catalonia	0.86 (0.75-0.92)	13 (16.1)	6 (11.3)
Madrid	0.81 (0.68-0.90)	15 (21.4)	7 (13.2)
Finland	0.47 (0.24-0.65)	9 (15.8)	3 (5.7)
Norway	0.40 (0.14-0.59)	15 (32.8)	3 (5.7)
Brazil	0.94 (0.90-0.96)	359 (22.6)	8 (15.1)
São Paulo	0.68 (0.42-0.83)	48 (30)	2 (3.8)
Rio Grande do Sul	0.89 (0.84-0.94)	19 (20)	0 (0)
Bahia	0.85 (0.77-0.92)	63 (24.3)	5 (9.4)

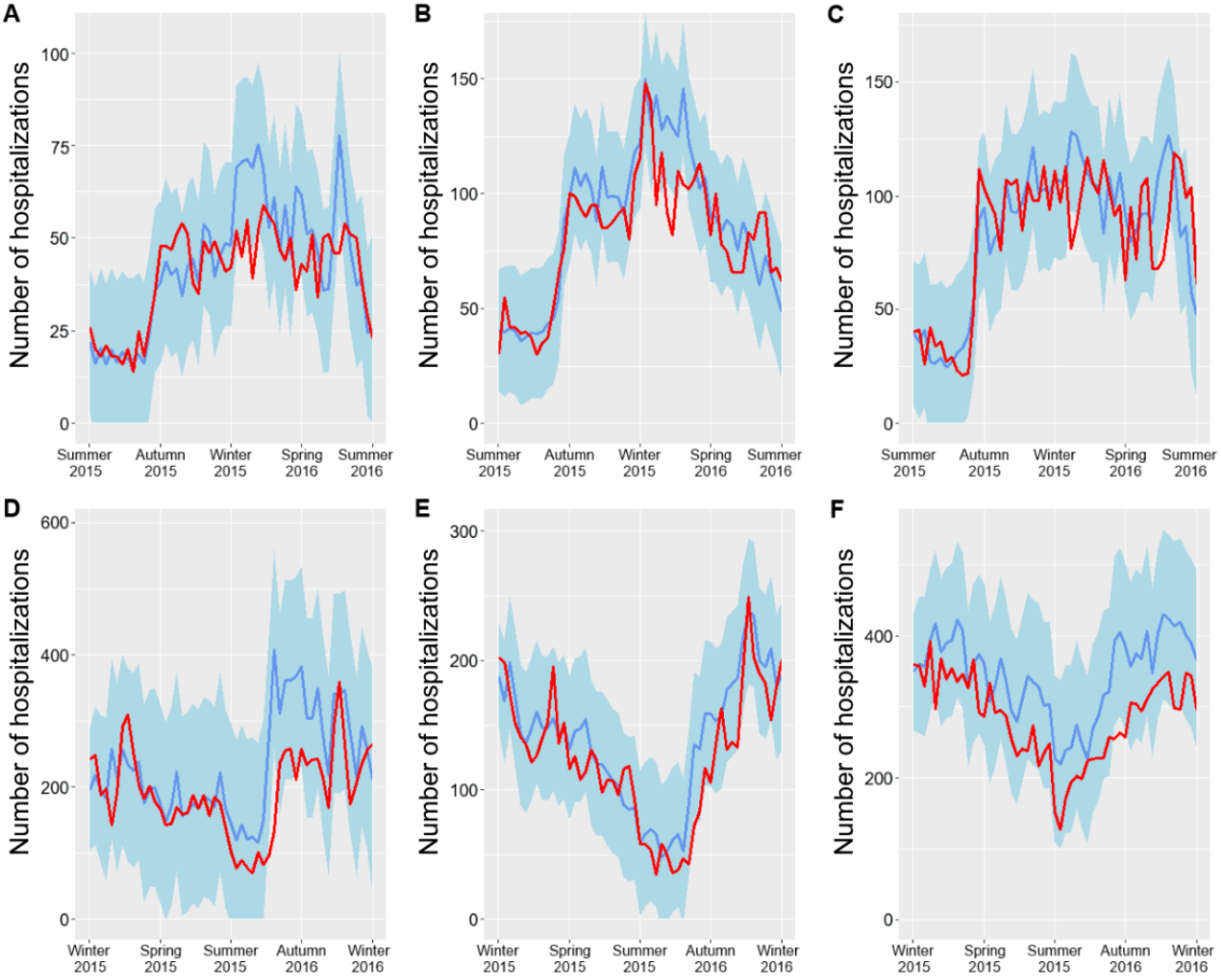
XSL•FO
RenderX

Figure 4. Predicted and observed number of asthma hospitalizations for 1 year in Portugal, Spain, Finland, Norway, and Brazil. Predicted hospitalizations were estimated based on previous hospitalizations and on Google Trends data for the *pseudo-influenza syndrome* topic.



XSL•FO
RenderX

Figure 5. Predicted (blue; 95% CI in light blue) and observed (red) number of asthma hospitalizations for 1 year in the Spanish Autonomous Communities of Andalusia (A), Catalonia (B), and Madrid (C), as well as in the Brazilian States of São Paulo (D), Rio Grande do Sul (E), and Bahia (F).



Discussion

Principal Findings

In this study, we observed moderate-to-strong correlations between rhinovirus-related GT and asthma hospitalization data in 5 different countries (Portugal, Spain, Finland, Norway, and Brazil). In addition, based on previous admission patterns and rhinovirus GT, we built seasonal ARIMA models with good capacity to forecast asthma hospitalizations.

Although the overall observed correlations were moderate to strong, and the forecast models showed good performance, between-country differences should be highlighted. Overall, lower correlations between rhinovirus-related GT and hospitalization data were observed in Finland and Norway. In these 2 countries, the analysis of our data indicates that asthma hospitalizations have markedly decreased throughout the years. Such a decrease may be partly explained by a focus on the early detection and prevention of exacerbations. Such goals are indeed stated in the Finnish Asthma Programme 1994-2004 and in the Finnish Allergy Programme 2008-2018, which have both promoted guided self-management (ie, patients' identification of causes and proactive prevention of exacerbations) as the primary form of treatment [16,17]. This better overall asthma control (with a more effective prevention of respiratory infection–related exacerbations) may in part be explained by the lower correlations between common cold–related GT and asthma hospitalizations. On the other hand, differences between Finland and Norway may be explained by the more frequent fluctuations in Norwegian asthma hospitalizations (even within the same season), which can be explained by geographic or climatic reasons: Norway has a long coast along several seas in the Atlantic and Arctic with more than five Köppen climate zones, whereas Finland only lies in the Baltic Sea with three Köppen climate zones. In fact, for Norway, a stronger cross-correlation (0.46) was observed when performing analyses based on 15-day average values instead of weekly values. These frequent fluctuations also explain why, despite the moderate correlations, the observed number of Norwegian hospitalizations fell within the forecasted 95% CIs in 49-51 out of 52 weeks (June 2015-June 2016).

On the other hand, within-country differences should also be considered. For example, in Spain, the strongest correlations between observed and predicted asthma hospitalizations were related to the autonomous community of Catalonia. The fact that correlations were weaker in other areas may be partly explained by the seasonal asthma hospitalization peaks observed

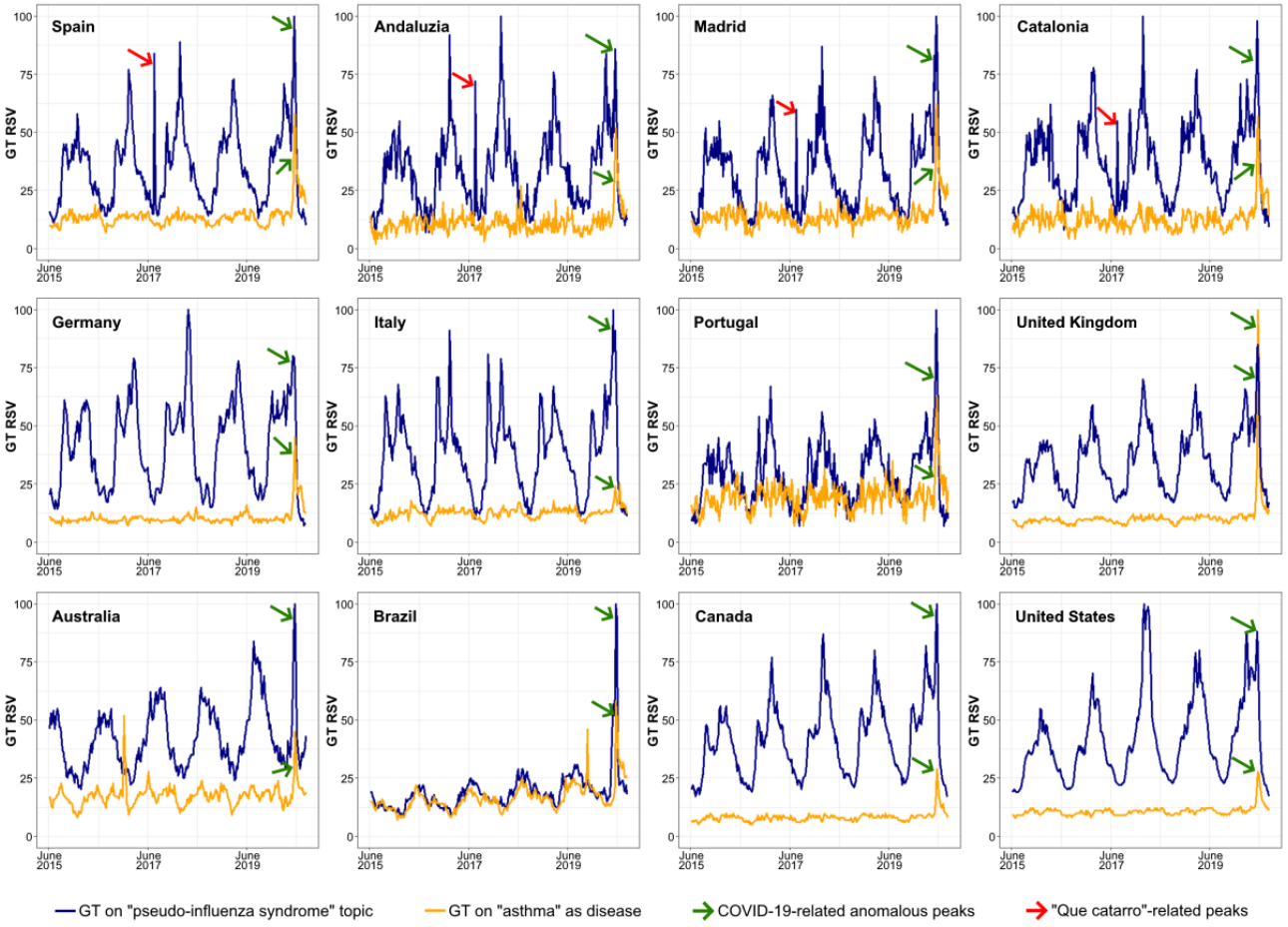
XSL•FO
RenderX

in Madrid and Andalusia, which had no correspondence with rhinovirus-related GT data. In fact, such peaks correlate with olive pollen peaks, and olive pollen concentrations are much higher in Madrid and Andalusia than in Catalonia [18]. Differences in pollen exposure may also partly explain the worse results observed in Portugal (when compared with Spain), where olive trees are abundant throughout all inland regions. In Brazil, the best results were consistently observed for Rio Grande do Sul, a state with only two Köppen climate zones, compared with seven in São Paulo and nine in Bahia. Therefore, although rhinovirus-related GT, along with previous hospitalization patterns, may help to forecast trends in asthma hospitalizations, improved accuracy is to be expected if other variables are also taken into account. Such variables may include pollen exposure, pollution levels, and even meteorological variables. This is in line with the multifactorial nature of asthma exacerbations, in which viral infections play a major role along with other environmental factors.

Strengths and Limitations

This study had several important limitations. First, GT provides information on users' search behaviors, which reflect not only the true epidemiological situation of a disease or condition but also other factors, such as media coverage or users' interest or curiosity [6]. Such factors are particularly evident when assessing anomalous peaks in pseudo-influenza GT data over the last 5 years (2015-2020). For example, in all regions of Spain (Figure 6; red arrows), but not in other countries, there was a 1-day peak at the end of June 2017. This peak was probably related to the TV cartoon episode *Qué catarro*, which was aired at that time across Spain [19,20] (*catarro* is a Spanish or Castilian word for *common cold*). In another example, a larger anomalous pseudo-influenza GT data peak was found in several countries during the COVID-19 outbreak in 2020 (Figure 6; green arrows). Previous studies have shown that GT can be highly influenced by media attention. For example, for several countries in Europe and America, search peaks on anosmia or ageusia had a better correlation with media coverage than with the epidemiological situation of COVID-19 [21]. Media coverage also appears to have played a decisive role in driving search peaks for asthma during the COVID-19 pandemic [11]. Although, at first sight, we could believe that media coverage does not particularly bias GT data in this study (as the common cold is a frequent and mild condition that does not drive media attention), the previously discussed Spanish anomalous data peak suggests that this factor should not be discarded.

Figure 6. *Pseudo-influenza syndrome* Google Trends data for the period 2015-2020, with anomalous peaks evidenced (anomalous peaks associated with COVID-19: green arrows; peaks associated with transmission of the cartoon *Que catarro* on Spanish television in June 2017: red arrows). GT: Google Trends; RSV: relative search volume.



Additional limitations include the absence of GT data on the actual number of performed searches, with only normalized data provided (ie, data expressed as a percentage of the search volume in the period during which the term or expression

XSL•FO
RenderX

gathered the most attention). In states or regions with a low volume of searches, the quality of the data provided by GT is not sufficient for analysis. As a result, we were not able to assess all Brazilian states (not even the most populous Brazilian state in every region) nor all Spanish autonomous communities. Nevertheless, we assessed the most populous Brazilian state in three of the five official geographical regions, as well as the three most populous Spanish autonomous communities. Another limitation concerns the secular increase in the use of the internet (and therefore of Google searching) between 2012 and 2016, particularly with the generalization of smartphone use. The share of searches using Google may also vary during this period. In Brazil, Google was responsible for 85% of the searches performed in 2013, but this percentage increased to over 95% in 2015 [22,23]. To control these secular trends, we applied time series analysis methods, removing the estimated trend components for both GT and hospitalization data.

Finally, in this study, we solely considered asthma hospitalizations, even though only a minority of exacerbations resulted in hospitalization. Although trends in hospitalizations probably mirror those in exacerbations, an assessment of both hospitalizations and emergency department visits (whose data were not currently available for analysis) would probably more accurately reflect trends in exacerbations. Furthermore, the generalizability of our forecast models is limited, as such models require previous hospitalization data, which may not be easily available in all countries.

This study had several strengths. In particular, we assessed 5 different countries (4 in Europe and 1 in South America) using nationwide data for a period of 5 years. For 2 of these countries, we considered regional differences to assess state-level data. In addition, rhinovirus-related GT data were retrieved by two different strategies—GT data on the *pseudo-influenza syndrome* topic and on search terms regarding the common cold—which obtained comparable results. Finally, for data analysis, we applied methods to remove secular trends of GT data and hospitalizations and built seasonal ARIMA models with predictions of hospitalization trends and frequencies that had similar results.

Conclusions

In this study, we found that rhinovirus-related GT data correlated well with asthma hospitalizations in Portugal, Spain, and Brazil (with moderate correlations observed for Finland and Norway). In addition, such GT data, along with previous admission patterns, were able to reasonably forecast asthma hospitalizations in the examined countries. Although these results suggest that rhinovirus-related GT data may be helpful when building models to predict asthma hospitalizations, future studies should explore the design of more complex models, taking environmental variables into account and possibly assessing exacerbations or other outcome variables.

Acknowledgments

The authors wish to thank the Portuguese Ministry of Health for providing access to the hospitalization data managed by the Portuguese Central Health System Administration (Administração Central do Sistema de Saúde). Data from the Norwegian Patient Registry were used in this study. The interpretation and reporting of these data are the sole responsibility of the authors, and no endorsement by the Norwegian Patient Registry is intended nor should it be inferred. Article Processing Charges were funded by National Funds through FCT-Fundação para a Ciência e a Tecnologia, IP, within CINTESIS, R&D Unit (reference UIDB/4255/2020).

Conflicts of Interest

None declared.

References

1. Bahadori K, Doyle-Waters MM, Marra C, Lynd L, Alasaly K, Swiston J, et al. Economic burden of asthma: a systematic review. BMC Pulm Med 2009 May 19;9:24 [FREE Full text] [doi: 10.1186/1471-2466-9-24] [Medline: 19454036]
2. Busse WW, Lemanske RF, Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. Lancet 2010 Sep 04;376(9743):826-834 [FREE Full text] [doi: 10.1016/S0140-6736(10)61380-3] [Medline: 20816549]
3. Nicholson KG, Kent J, Ireland DC. Respiratory viruses and exacerbations of asthma in adults. Br Med J 1993 Oct 16;307(6910):982-986 [FREE Full text] [doi: 10.1136/bmj.307.6910.982] [Medline: 8241910]
4. Johnston SL, Pattemore PK, Sanderson G, Smith S, Lampe F, Josephs L, et al. Community study of role of viral infections in exacerbations of asthma in 9-11 year old children. Br Med J 1995 May 13;310(6989):1225-1229 [FREE Full text] [doi: 10.1136/bmj.310.6989.1225] [Medline: 7767192]
5. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
6. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

XSL•FO
RenderX

7. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating smart health in the US health care system: infodemiology study of asthma monitoring in the Google era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24 [FREE Full text] [doi: 10.2196/publichealth.8726] [Medline: 29530839]
8. Osborne NJ, Alcock I, Wheeler BW, Hajat S, Sarran C, Clewlow Y, et al. Pollen exposure and hospitalization due to asthma exacerbations: daily time series in a European city. Int J Biometeorol 2017 Oct;61(10):1837-1848 [FREE Full text] [doi: 10.1007/s00484-017-1369-2] [Medline: 28500390]
9. Bousquet J, Onorato GL, Oliver G, Basagana X, Annesi-Maesano I, Arnavielhe S, et al. Google Trends and pollen concentrations in allergy and airway diseases in France. Allergy 2019 Oct;74(10):1910-1919. [doi: 10.1111/all.13804] [Medline: 30942904]
10. Bousquet J, O'Hehir RE, Anto JM, D'Amato G, Mösges R, Hellings PW, et al. Assessment of thunderstorm-induced asthma using Google Trends. J Allergy Clin Immunol 2017 Sep;140(3):891-893. [doi: 10.1016/j.jaci.2017.04.042] [Medline: 28579375]
11. Sousa-Pinto B, Heffler E, Antó A, Czarlewski W, Bedbrook A, Gemicioglu B, et al. Anomalous asthma and chronic obstructive pulmonary disease Google Trends patterns during the COVID-19 pandemic. Clin Transl Allergy 2020 Nov 02;10(1):47 [FREE Full text] [doi: 10.1186/s13601-020-00352-9] [Medline: 33292529]
12. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
13. Bousquet J, Agache I, Anto JM, Bergmann KC, Bachert C, Annesi-Maesano I, et al. Google Trends terms reporting rhinitis and related topics differ in European countries. Allergy 2017 Aug;72(8):1261-1266. [doi: 10.1111/all.13137] [Medline: 28140507]
14. Bousquet J, Agache I, Berger U, Bergmann K, Besancenot J, Bousquet PJ, et al. Differences in reporting the ragweed pollen season using Google Trends across 15 countries. Int Arch Allergy Immunol 2018;176(3-4):181-188 [FREE Full text] [doi: 10.1159/000488391] [Medline: 29742519]
15. Song X, Xiao J, Deng J, Kang Q, Zhang Y, Xu J. Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011. Medicine (Baltimore) 2016 Jun;95(26):e3929 [FREE Full text] [doi: 10.1097/MD.0000000000003929] [Medline: 27367989]
16. Haahtela T, Tuomisto LE, Pietinalho A, Klaukka T, Erhola M, Kaila M, et al. A 10 year asthma programme in Finland: major change for the better. Thorax 2006 Aug;61(8):663-670 [FREE Full text] [doi: 10.1136/thx.2005.055699] [Medline: 16877690]
17. Haahtela T, Valovirta E, Bousquet J, Mäkelä M, Allergy Programme Steering Group. The Finnish Allergy Programme 2008-2018 works. Eur Respir J 2017 Jun;49(6):1700470 [FREE Full text] [doi: 10.1183/13993003.00470-2017] [Medline: 28642312]
18. SILAM v.5.7: System for Integrated modeLling of Atmospheric coMposition. Finnish Meteorological Institute. 2020. URL: http://silam.fmi.fi/ [accessed 2020-06-15]
19. Programación Clan TVE. El Mundo. 2017. URL: https://web.archive.org/web/20170630042653/http://www.elmundo.es/television/programacion-tv/clan-tve.html [accessed 2020-11-01]
20. Tutu: Qué catarro CLAN TVE. 2017. URL: https://www.rtve.es/infantil/ [accessed 2020-11-01]
21. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google Trends data: infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
22. 85% de participação, Google é o buscardor mais usado no Brasil [With 85% of share, Google is the most used search engine in Brazil]. G1 Globo. 2013. URL: http://g1.globo.com/tecnologia/noticia/2013/07/com-85-de-participacao-google-e-o-buscador-mais-usado-no-brasil.html [accessed 2020-06-15]
23. Rocha C. Confira o ranking dos buscadores no Brasil [Check out the ranking of search engines in Brazil]. Olhar Digital. 2015. URL: https://web.archive.org/web/20201023114542/https://olhardigital.com.br/noticia/ranking-dos-buscadores-no-brasil/46539 [accessed 2020-06-15]

Abbreviations

ARIMA: autoregressive integrated moving average
GT: Google Trends

XSL•FO
RenderX

Edited by R Kukafka; submitted 08.01.21; peer-reviewed by A Mavragani, P Satpathy, S Huynh; comments to author 15.02.21; revised version received 04.03.21; accepted 24.03.21; published 06.07.21.

Please cite as:
Sousa-Pinto B, Halonen JI, Antó A, Jormanainen V, Czarlewski W, Bedbrook A, Papadopoulos NG, Freitas A, Haahtela T, Antó JM, Fonseca JA, Bousquet J
Prediction of Asthma Hospitalizations for the Common Cold Using Google Trends: Infodemiology Study
J Med Internet Res 2021;23(7):e27044
URL: https://www.jmir.org/2021/7/e27044
doi:10.2196/27044
PMID:34255692

©Bernardo Sousa-Pinto, Jaana I Halonen, Aram Antó, Vesa Jormanainen, Wienczyslawa Czarlewski, Anna Bedbrook, Nikolaos G Papadopoulos, Alberto Freitas, Tari Haahtela, Josep M Antó, João Almeida Fonseca, Jean Bousquet. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing Patient Perceptions and Experiences of Paracetamol in France: Infodemiology Study Using Social Media Data Mining

Stéphane Schück[1], MSc, MD; Avesta Roustamal[1], MSc; Anaïs Gedik[1], MSc; Paméla Voillot[1], MSc; Pierre Foulquié[1], MSc; Catherine Penfornis[2], MD; Bernard Job[2], MD

[1]Kap Code, Paris, France
[2]Sanofi, Paris, France

Corresponding Author:
Avesta Roustamal, MSc
Kap Code
28 Rue d'Enghien
Paris,
France
Phone: 33 9 72 60 57 64
Email: avesta.roustamal@kapcode.fr

Abstract

Background: Individuals frequently turning to social media to discuss medical conditions and medication, sharing their experiences and information and asking questions among themselves. These online discussions can provide valuable insights into individual perceptions of medical treatment, and increasingly, studies are focusing on the potential use of this information to improve health care management.

Objective: The objective of this infodemiology study was to identify social media posts mentioning paracetamol-containing products to develop a better understanding of patients' opinions and perceptions of the drug.

Methods: Posts between January 2003 and March 2019 containing at least one mention of paracetamol were extracted from 18 French forums in May 2019 with the use of the Detec't (Kap Code) web crawler. Posts were then analyzed using the automated Detec't tool, which uses machine learning and text mining methods to inspect social media posts and extract relevant content. Posts were classified into groups: Paracetamol Only, Paracetamol and Opioids, Paracetamol and Others, and the Aggregate group.

Results: Overall, 44,283 posts were analyzed from 20,883 different users. Post volume over the study period showed a peak in activity between 2009 and 2012, as well as a spike in 2017 in the Aggregate group. The number of posts tended to be higher during winter each year. Posts were made predominantly by women (14,897/20,883, 71.34%), with 12.00% (2507/20,883) made by men and 16.67% (3479/20,883) by individuals of unknown gender. The mean age of web users was 39 (SD 19) years. In the Aggregate group, pain was the most common medical concept discussed (22,257/37,863, 58.78%), and paracetamol risk was the most common discussion topic, addressed in 20.36% (8902/43,725) of posts. Doliprane was the most common medication mentioned (14,058/44,283, 31.74%) within the Aggregate group, and tramadol was the most commonly mentioned drug in combination with paracetamol in the Aggregate group (1038/19,587, 5.30%). The most common unapproved indication mentioned within the Paracetamol Only group was fatigue (190/616, with 16.32% positive for an unapproved indication), with reference to dependence made by 1.61% (136/8470) of the web users, accounting for 1.33% (171/12,843) of the posts in the Paracetamol Only group. Dependence mentions in the Paracetamol and Opioids group were provided by 6.94% (248/3576) of web users, accounting for 5.44% (342/6281) of total posts. Reference to overdose was made by 245 web users across 291 posts within the Paracetamol Only group. The most common potential adverse event detected was nausea (306/12843, 2.38%) within the Paracetamol Only group.

Conclusions: The use of social media mining with the Detec't tool provided valuable information on the perceptions and understanding of the web users, highlighting areas where providing more information for the general public on paracetamol, as well as other medications, may be of benefit.

(J Med Internet Res 2021;23(7):e25049) doi:10.2196/25049

KEYWORDS

analgesic use; data mining; infodemiology; paracetamol; pharmacovigilance; social media; patient perception

XSL•FO
RenderX

Introduction

Background

Over-the-counter (OTC) medications are generally effective and associated with a well-documented safety profile when used as directed, making them convenient for managing ailments such as mild pain when medical consultation is not required [1]. The responsibility for the proper use of these medications falls on the individual using them, who may be guided by a pharmacist. While some information is available about an individual's decision-making process when selecting medication, including demographic and social factors [2], more remains to be learned about issues that may affect an individual's approach to taking OTC medications and whether additional information and guidance can further improve the safety and effectiveness of OTC medications. Areas of interest include symptom recognition (Do individuals understand their conditions?), self-selection (Why do they seek relief for their symptoms on their own?), active ingredients (Are the active ingredients known?), dosing (Do individuals distinguish between prescription regimens and use-as-needed instructions?), concomitant use warnings (Do individuals know the ingredients they need to heed because of warnings?), and when to stop the medication (Do individuals heed the maximum daily dose or duration of use warnings?).

Paracetamol is a well-established medication for pain management with a good safety profile when used as recommended, but increasingly, reports of inadequate use, including medication errors and misuse of the drug, have arisen [1,2]. In a study of OTC products containing paracetamol, 24% of adults confirmed they would take more than the recommended maximum dose within a 24-hour period; more than 20% of self-treating people struggled with dosing timing, such as taking another dose too soon; and 46% used more than one product with the same active ingredients [3]. Another study found many individuals do not routinely examine product label information for OTC pain relief, and more than 50% are unaware of the active ingredient [4]. The French National Agency for Medicines and Health Products Safety (ANSM) has therefore launched a campaign to raise awareness of the toxicity risks associated with the misuse of paracetamol [5].

Over the last decade, people have been increasingly using social media, including health forums, to communicate about and further understand all aspects of disease and treatment [6,7]. Compared with traditional medical data, social media can offer valuable insights into hundreds of thousands of individuals' treatment management, including their opinions and concerns [8-10]. Topic mining in social media can document all the above aspects related to an individual's experience with paracetamol and may also provide an adjunct to pharmacovigilance activities [11]. Different methods of social media data mining and topic modeling have been under investigation for use in pharmacovigilance studies to assist with postmarketing surveillance of drugs [12,13], and both the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) are investigating the possibility of social media as a new data source to strengthen their activities regarding drug safety [14,15]. Detec't (Kap Code, Paris, France) is an automated tool that uses artificial intelligence and text mining methods to analyze social media posts and extract relevant content. The Detec't tool has previously been used to mine patient narratives from popular French forums for details such as potential adverse events (PAEs) and noncompliance, with a high level of specificity [11,13,16,17].

Objective

Using the Detec't tool, this infodemiology study investigated the opinions expressed by individuals on social media by identifying posts mentioning paracetamol-containing products.

Methods

Data Extraction

This was a retrospective infodemiology study assessing the contents of posts in social media about paracetamol products. A search was conducted in May 2019 to identify all posts made between January 2003 and March 2019, posts containing at least one keyword relating to paracetamol products were extracted from 18 general and specialized French social media channels (Multimedia Appendix 1) using the Detec't web crawler as previously described [11,16]. Web scraping of the messages was performed depending on the HTML structure of each forum. Posts containing at least one keyword (Multimedia Appendix 2) were automatically retrieved with all the associated metadata, deidentified, and cleaned (signature and quote withdrawal).

End Points

The primary end point was to identify the most frequent discussion themes in social media channels that mention paracetamol. The secondary end point was to identify PAEs for aggregated data analysis, including mentions of incorrect drug use.

Preprocessing

The analysis corpus was cleaned after the removal of unrelated messages and posts in languages other than French. The posts were classified into Paracetamol Only, Paracetamol and Opioids, Paracetamol and Others, and Aggregate groups. The Paracetamol Only group included any post with only the predetermined 20 words associated with paracetamol; Paracetamol and Opioids included any drugs containing paracetamol and an opioid such as codeine or tramadol; Paracetamol and Others included drugs containing paracetamol and another nonopioid molecule, such as vitamin C; and the Aggregate group encompassed all 3 categories and contained all posts mentioned. Duplicate posts were removed; however, posts could be incorporated into more than one group if keywords relevant to multiple groups were identified.

Processing and Statistical Analysis

Monitoring Algorithm

Postactivity volume information was identified with the use of a Markov chain–based algorithm, which clustered the data according to the time of posting [18,19]. The posts were separated into weak, moderate, or high activity; categories were

XSL•FO
RenderX

defined not only regarding post volumes but also temporal dynamics, as permitted by Markov models. This allowed for the identification of unusual activity periods; for example, a marked increase in posts could correspond with a release of new safety-related information. A time series was created for the volume of posts per month, and data were corrected for periodic seasonality using locally estimated scatterplot smoothing (LOESS) regression on each seasonal subseries [20]. Estimation of the Hidden Markov Model (HMM) was performed using the expectation-maximization algorithm. The initial number of centroids was arbitrarily set to 3 in order to provide a simple categorization of posts' volume observations. Observations were then clustered using a K-means clustering on the series trend and values, as well as the output of the HMM. Finally, smoothing of the clustering output was performed to obtain sequences of activity types.

Age and Gender

The gender of users was determined with the identification of specific expressions within the message (gendered past participles, adjectives, and names), as well as the application of a support vector machine (SVM) separating posts into female, male, or unknown. The model for gender identification was a linear SVM with a cost of 1. Age categories were also determined based on the use of specific expressions within the message.

Medical Concepts

The corpus of paracetamol posts that contained medical concepts was identified with the use of the Medical Dictionary for Regulatory Activities (MedDRA) version 15.0 and enriched with a web user vocabulary, as described elsewhere [16].

Discussion Topics

A topic model was applied to identify the themes addressed within the messages. These models are based on the hypothesis that each document in the corpus corresponds to a distribution of several topics. The modeled topics are probability distributions over the tokens (words or sequences of several adjacent words) found within the corpus. No prior assumption was made about the nature of topics present in the corpus. These models have been used previously to analyze health-related topics within web forums [21,22].

For this study, the correlated topic model based on the latent Dirichlet allocation was used [23]. The modeling of the studied corpus went through different preprocessing and cleaning steps so that the topic model could be applied. The model was estimated using a variational expectation-maximization algorithm. Topics, being probability distributions over tokens of the corpus of study, could be characterized by the highest per-topic probability tokens. Evaluating these probabilities through term-frequency inverse document frequency (TF-IDF) weighting allows the allocation of a higher importance to the topic specific tokens. In this case, the per-topic probability of a token was weighted by the inverse of the probabilities of this token in other topics. For each topic, tokens were ranked from highest- to lowest-weighted probabilities as per the TF-IDF value of their probability in this topic [24]. The first 15 tokens were designated as the set of characteristic tokens and used to name the topic. The analysis was performed using the structural topic model package [25] with R environment (version 3.5.2, R Foundation for Statistical Computing).

Top 10 Treatments

The top 10 treatments include the first 10 paracetamol-containing brands discussed most within the messages. The products mentioned were identified, and the occurrence of the medications in the posts was counted.

Drug-Intake Algorithm

A drug intake algorithm was used to determine whether each drug identified in the post had been taken by the web user. The messages were separated into individual sentences before specific variables were applied. Three distinct variables were used: lexical features (eg, intake and nonintake lexical fields), stylistic features (eg, proportion of exclamation marks, proportion of used pronouns), and syntactic structure features (eg, length of sentences, drug position). Analyzed content was scaled to avoid bias in the direction of the user who posted richer content (eg, number of posts or diversity of content). A random forest algorithm was applied to predict whether drug intake was expressed or not at the sentence level. Each sentence prediction was aggregated per post in order to have a global intake prediction per message.

Simultaneous Drug Consumption

Simultaneous drug consumption data were identified by assessing, per web user, the consumption of a paracetamol product (as per the list of predefined paracetamol-associated keywords) at the same time as another drug, detected with the use of the Detec't medical product list (which contains approximately 2500 molecules and drugs). The drug intake algorithm was applied only to posts that mentioned paracetamol consumption, and the named-entity recognition technique allowed for the identification and labeling of the other drugs mentioned within the message.

Potential Adverse Events

Messages that included mention of drug intake were identified and then assessed for medical concepts (as determined with MedDRA). An SVM classifier, with a weighted radial SVM with cost 100 and gamma parameter 0.1, trained on an annotated gold standard [16], was then used to assess the messages to determine whether the post mentioning a paracetamol-containing product included a PAE. A machine learning model was used to train the SVM classifier; this used several clustering parameters, including the distance between the concept and the closest drug mention, the length of the message, and the number of times this concept was identified within the message. This clustering method is described elsewhere [16]. A manual review was performed by experts on one-third of the posts related to PAEs.

Unapproved Indications, Dependence, and Overdose

The analysis on incorrect use encompassed unapproved indications, dependence, and overdose. In order to assess incorrect drug use within adults only, data from use in children (aged younger than 15 years) were eliminated.

XSL•FO
RenderX

Unapproved Indications

Within each post expressing drug intake and medical concepts, stated drug use was assessed, manually filtered, and compared with the approved indications found within the product Summary of Product Characteristics (SmPC). These SmPC terms (from MedDRA) were used for each category of drug. Unapproved indications were defined as situations where a medical product was intentionally used for a purpose not in accordance with the terms of the marketing authorization. Posts containing medical concepts that were approved indications were excluded from the analysis corpus. The remainder were reviewed manually and for each unapproved indication selected by categories; 10% of messages were randomly sampled and annotated manually. The percentage shown represents the percentage of posts detected for the unapproved indication that were confirmed manually to have an unapproved indication.

Dependence

Dependence was determined by calculating a score through the BM25 [26] algorithm comparing a message and a reference pattern. The reference pattern was composed from a dependency lexical field, obtained from a sample of messages. The dependency lexical field contained words and expressions linked to drug dependency, which were manually reviewed by a group of experts.

All messages were divided into sentences containing the drug name and at least one of the words from the dependency lexical field. Each sentence within a message was then compared to the pattern reference using the BM25 score. The general score for each message was calculated by summation of each message. Higher scores revealed a similarity between the reference pattern and the message (eg, the higher the score, the more the sentence expressed dependency on a drug). To classify a message as discussing dependency, a similarity score threshold was chosen. Within this study, if a message had a similarity score equal to or greater than 40, it was identified as a message expressing dependency.

Overdose

The paracetamol overdose algorithm was based on linguistic rules used to construct an algorithm based on pattern matching. The algorithm highlighted overdose expressions by matching expressions constructed on linguistic and syntactic rules as “I[PRONOUN] take [VERB] 10[DIGIT]gr [DOSAGE EXPRESSION]” or lexical field related to an overdose as “overdose,” “addicted to paracetamol.” These expressions were obtained from manual reviews provided by a group of experts. The paracetamol dose contained in each unit of drugs (eg, pill or sachet) was reviewed manually and the quantity converted into units or grams. In the event of a discrepancy over the amount, a higher dose was assumed.

The paracetamol dose identified in the post was converted to a per-day dose by the algorithm and compared with the paracetamol daily dose threshold [27]. If the dosage expressed in a post was greater than the daily recommended dose of paracetamol, or if the post contained an expression from the lexical field of overdose, the post was classified as expressing an overdose. The overdose was calculated as a percent above the daily dose of the product in question. The algorithm was also able to detect lexical field words related to overdose. The number of users expressing a drug overdose was calculated by adding the number of users detected with the digits and dosage expressions above the daily dose, as well as the number of users detected with overdose lexical field expressions. Each user was counted once.

Ethical Considerations

Data collection and treatment followed the European Union General Data Protection Regulation. A privacy-by-design approach was adopted as retrieved posts were anonymized before being stored in the analysis corpus. Furthermore, all presented results were aggregated.

Results

In May 2019, a search was conducted across 18 French forums (Multimedia Appendix 1) for posts made between January 2003 and March 2019. Figure 1 shows the distribution of posts associated with each category.

The Aggregate group contained 44,283 posts from 20,883 different users. Comparatively, the Paracetamol Only group had 33,196 messages (74.96% of total corpus) from 17,070 users (81.74% of total users); the Paracetamol and Opioids group had 14,733 (33.27%) messages, accounting for 6838 users (33.27%); and the Paracetamol and Others group had 1224 messages (2.76%) from 828 users (3.69%).

Figure 2 shows the variable activity seen in the post volume over the study period for the Aggregate group. A peak in postvolume activity was seen between 2009 and 2012, with a spike in activity seen in the middle of 2017. The number of posts tended to be higher during the winter and lower during the summer. Post activity for the individual groups is presented in Multimedia Appendices 3-5. The sociodemographic characteristics analysis shows that the posts were predominantly from women and the mean age of web users was 39 (SD 19) years (Table 1).

For the primary end point, pain was the most common medical concept discussed within the Paracetamol Only group with 15,310 posts (57.20%), as well as the Paracetamol and Opioids group with 6710 posts (65.41%; Table 2). Within the Paracetamol and Others group, nasopharyngitis was listed as the most common medical concept with 251 posts (30.0%), followed by pain (237 posts; 28.3%). Table 3 shows discussion topics identified across all 4 main groups, providing more specific information on the common threads web users were discussing. Paracetamol risk was the most common topic for the Aggregate group, addressed in 8902 posts (20.36%), followed by depression and suicide attempts found in 8063 posts (18.44%). Drug intake in everyday life, however, was the most common discussion topic for both the Paracetamol Only group (10,949 posts, 33.37%) and the Paracetamol and Opioids group (4235 posts, 29.10%). For the Paracetamol and Others group, cold medicine was the most common discussion point, addressed in 404 posts (34.38%).

XSL•FO
RenderX

Figure 1. Categorization of posts containing paracetamol-related content. MedDRA: Medical Dictionary for Regulatory Activities; PAE: potential adverse event.

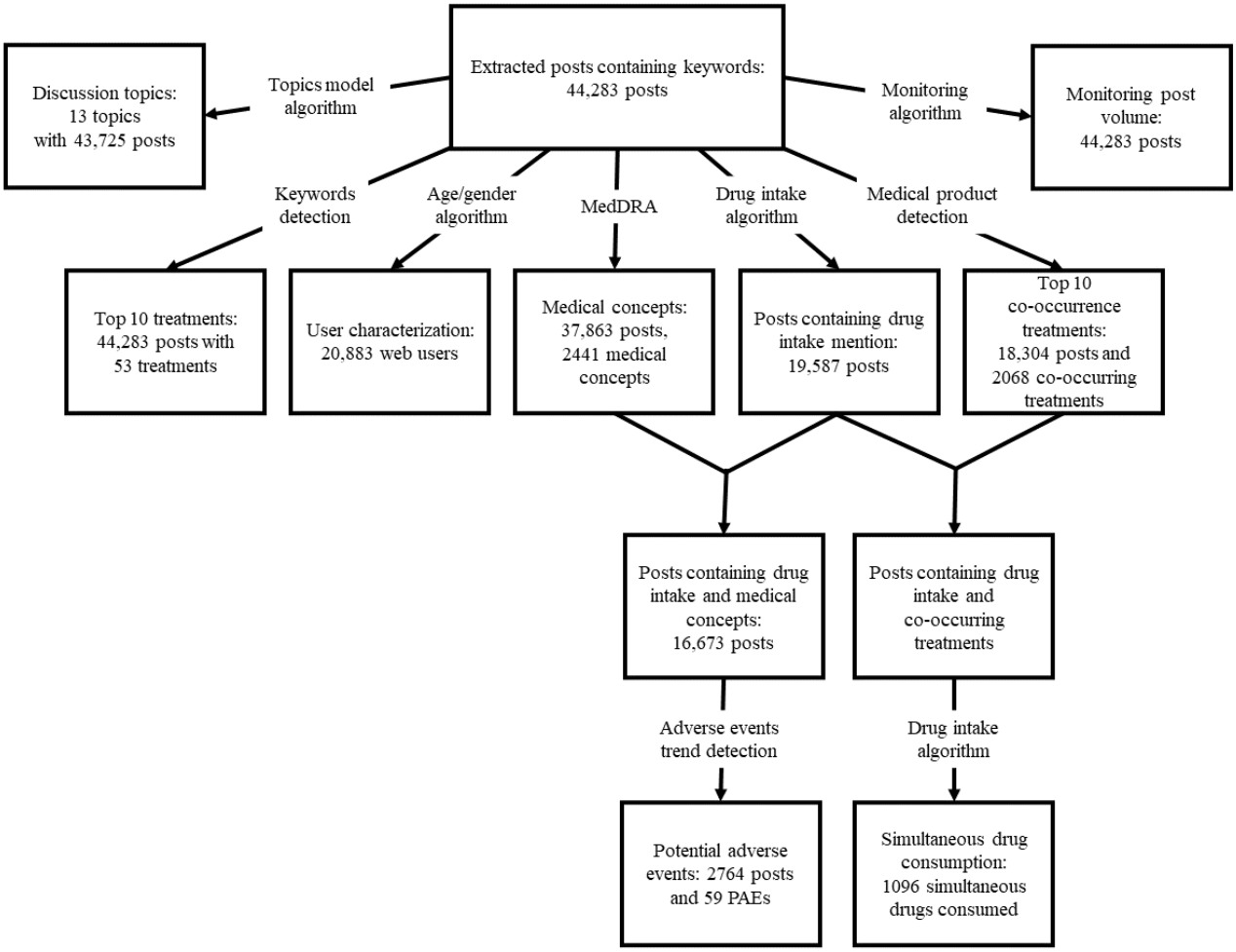
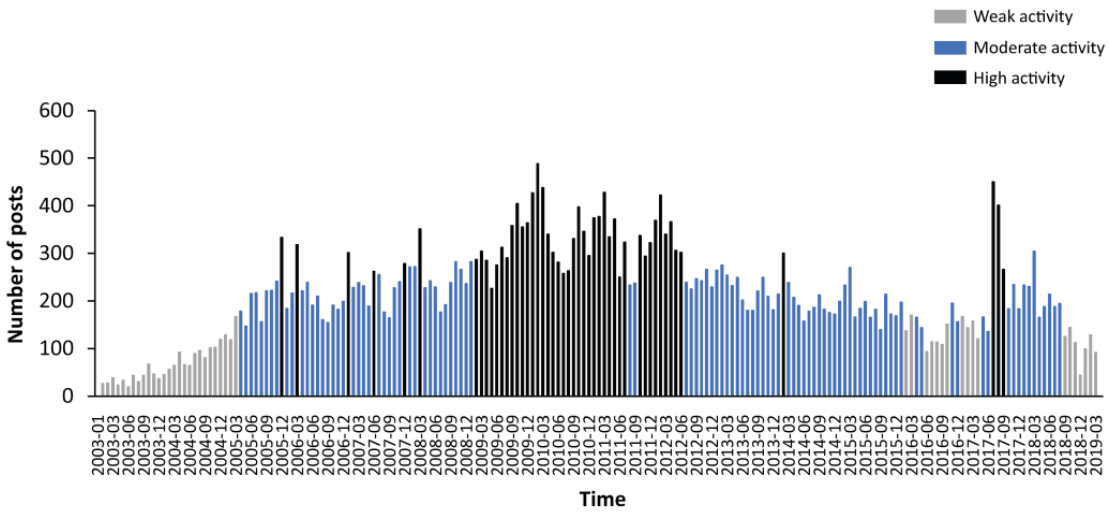


Figure 2. Volume of posts containing mention of paracetamol over time for the Aggregate group.



XSL•FO
RenderX

Table 1. Sociodemographic characteristics of web users.

Characteristic	Paracetamol only (n=17,070), n (%)	Paracetamol and opioids (n=6838), n (%)	Paracetamol and others (n=828), n (%)	Aggregate[a] (n=20,883), n (%)
Gender				
Women	11,999 (70.29)	4985 (72.90)	563 (68.00)	14,897 (71.34)
Men	1962 (11.49)	821 (12.01)	105 (13.04)	2507 (12.00)
Unknown	3109 (18.21)	1032 (15.09)	160 (19.32)	3479 (16.66)
Total	17,070 (100.00)	6838 (100.00)	828 (100.00)	20,883 (100.00)
Age group (years)				
0-20	1767 (10.35)	593 (8.67)	74 (8.93)	2104 (10.08)
21-30	4565 (26.74)	1596 (23.34)	242 (29.23)	5454 (26.12)
31-40	4590 (26.89)	1795 (26.25)	234 (28.26)	5609 (26.86)
41-50	1874 (10.98)	936 (13.69)	81 (9.78)	2409 (11.54)
51-60	926 (5.42)	479 (7.00)	49 (5.91)	1226 (5.87)
60+	1982 (11.61)	1050 (15.36)	62 (7.49)	2611 (12.50)
Unknown	1366 (8.00)	389 (5.69)	86 (10.39)	1470 (7.04)
Total	17,070 (100.00)	6838 (100.00)	828 (100.00)	20,883 (100.00)

[a]Aggregate users are the number of distinct users, and as such this column is not the sum of previous columns as a user may have posted in more than one other category.

Table 2. Top 10 medical concepts detected within the posts for each group.

Medical concept	Aggregate (n=37,863), n (%)	Paracetamol only (n=26,768), n (%)	Paracetamol and opioids (n=10,258), n (%)	Paracetamol and others (n=837), n (%)
Pain	22,257 (58.78)	15,310 (57.20)	6710 (65.41)	237 (28.32)
Fatigue	4279 (11.30)	2962 (11.07)	1244 (12.13)	73 (8.72)
Pyrexia	3564 (9.41)	3358 (12.54)	—[a]	61 (7.29)
Analgesic drug level	3193 (8.43)	1786 (6.67)	1295 (12.62)	112 (13.38)
Dependence	3177 (8.39)	1535 (5.73)	1586 (15.46)	56 (6.70)
Migraine	2840 (7.50)	1733 (6.47)	1032 (10.06)	75 (8.96)
Headache	2831 (7.48)	2091 (7.81)	672 (6.55)	68 (8.12)
Pregnancy	2322 (6.13)	1876 (7.01)	—	86 (10.27)
Adverse event	2022 (5.34)	—	733 (7.15)	—
Nausea	1898 (5.01)	1296 (4.84)	566 (5.51)	—
Vomiting	—	1335 (4.99)	—	—
Nasopharyngitis	—	—	—	251 (29.99)
Somnolence	—	—	—	55 (6.57)
Anxiety	—	—	559 (5.45)	—
Emotional distress	—	—	516 (5.03)	—

[a]Only the top 10 medical concepts are presented for each group. Instances where an entry is not listed does not mean that these medical concepts were not detected within the group.

XSL•FO
RenderX

Table 3. Most common discussion topics identified within the posts.

Topic	Posts, n (%)
Aggregate (n=43,725)	
Paracetamol risk	8902 (20.36)
Depression and suicide attempt	8063 (18.44)
Drug intake in everyday life	7342 (16.79)
Children's prescriptions	7320 (16.74)
Addiction	6554 (14.99)
Neuropathic pain	5604 (12.82)
Drug use during pregnancy	5481 (12.54)
Postsurgery pains	5456 (12.48)
Switch of medications	5357 (12.25)
Gynecology and paracetamol	5057 (11.57)
Posology and composition	4373 (10.00)
Migraine relief	3238 (7.41)
Alternative therapies	3021 (6.91)
Paracetamol only (n=32,807)	
Drug intake in everyday life	10,949 (33.37)
Gynecology and paracetamol	7085 (21.60)
Neuropathic pain	7025 (21.41)
Sharing information	5770 (17.59)
Addiction	5276 (16.08)
Children's prescriptions	4947 (15.08)
Sleep disorders	4675 (14.25)
Postsurgery pains	4284 (13.06)
Advice for pain relief	3612 (11.01)
Migraine relief	3590 (10.94)
Care pathway	3534 (10.77)
Paracetamol and opioids (n=14,617)	
Drug intake in everyday life	4253 (29.10)
Addiction	3427 (23.45)
Incorrect opioid use	3416 (23.37)
Migraine relief	3195 (21.86)
Neuropathic pain	3048 (20.85)
Depression and suicide attempt	3022 (20.67)
Chronic pain	2530 (17.31)
Postsurgery pain	2480 (16.97)
Acute pain treatment (gynecologic, toothaches)	2457 (16.81)
Posology and composition	282 (1.93)
Paracetamol and others (n=1175)	
Cold medicines	404 (34.38)
Ear, nose, and throat problems	371 (31.57)
Drug use during pregnancy	331 (28.17)
Adverse events	211 (17.96)

XSL•FO
RenderX

Topic	Posts, n (%)
Migraine relief	180 (15.32)
Dizziness sensations	138 (11.74)
Self-prescriptions	127 (10.81)
Therapeutic options	126 (10.72)
Posology and composition	71 (6.04)
Drugs during breastfeeding	26 (2.21)

The analysis displaying the top 10 treatments mentioned in the posts identified Doliprane as the most common medication with 14,058 posts (31.74%), followed by Dafalgan with 4812 posts (10.86%), and Ixprim with 4344 posts (9.80%). Assessing simultaneous drug consumption, tramadol was the drug most commonly mentioned in combination with paracetamol across all groups except Paracetamol and Others; in the Aggregate group, tramadol accounted for 1038 posts (5.30% of all posts suggesting a paracetamol drug intake), followed by ibuprofen (601 posts, 3.07%). In the Paracetamol and Others group, caffeine was most commonly consumed with paracetamol, accounting for 65 posts (14.04%).

Assessing posts containing possible incorrect use, as per the secondary end points, the algorithm for unapproved indication detection identified 190 posts associated with fatigue, the most common unapproved indication mentioned within the Paracetamol Only group. Following a manual review, 16.32% of these posts were found to be positive for unapproved indications, with the remainder of posts associated with fatigue describing fatigue as a symptom or an effect of the drug intake. Dependence was detected in 148 posts, of which 14.19% were manually validated for an unapproved indication. Within the Paracetamol and Opioids group, dependence was detected in 238 posts, of which 66.81% were manually validated for an unapproved indication. This was followed by fatigue, which was detected in 110 posts and manually validated in 9.10%.

Within the Paracetamol Only group, posts with reference to dependence (as identified with the use of the BM25 algorithm), were made by 1.61% of the web users, accounting for 1.33% of the posts. Comparatively, dependence mentions in the Paracetamol and Opioids group were provided by 6.94% of web users, accounting for 5.44% of total posts.

Reference to overdose was made by 245 web users across 291 posts within the Paracetamol Only group. The most referenced drug was Doliprane, mentioned by 88 web users. For the Paracetamol and Opioids group, reference to overdose was made by 128 web users across 177 posts, and Codoliprane was the most referenced drug, mentioned by 67 web users. Some posts referenced overdoses at up to 500% above the recommended daily dose.

The top 10 PAE trends, which include mention of drug intake, are displayed in Table 4. For the Paracetamol Only group, the most common event detected was nausea, mentioned in 2.38% of posts detected by the SVM classifier, followed by vomiting (2.27%) and hypersensitivity (1.24%). For the Paracetamol and Opioids group, the most common event was dependence (7.93% of posts), followed by somnolence (2.55%).

XSL•FO
RenderX

Table 4. Top 10 potential adverse event trends detected within the posts.

Potential adverse event trends	Posts, n (%)[a]
Paracetamol only (n=12,843)	
Nausea	306 (2.38)
Vomiting	292 (2.27)
Hypersensitivity	159 (1.24)
Substance abuse	141 (1.10)
Dizziness	107 (0.83)
Feeling abnormal	73 (0.57)
Malaise	68 (0.53)
Overdose	64 (0.50)
Chills	62 (0.48)
Feeling jittery	59 (0.46)
Paracetamol and opioids (n=6281)	
Dependence	498 (7.93)
Somnolence	160 (2.55)
Substance abuse	125 (1.99)
Insomnia	103 (1.64)
Dizziness	101 (1.61)
Substance-induced psychotic disorder	74 (1.18)
Withdrawal syndrome	71 (1.13)
Feeling abnormal	56 (0.89)
Constipation	47 (0.75)
Euphoric mood	40 (0.64)

[a]Percentage does not represent the frequency of occurrence in treated patients, but rather the percentage of social media posts found within the drug intake and medical concepts cohorts, in which mention of a given potential adverse event was detected with the adverse events detection algorithm.

Discussion

Principal Findings

This study aimed to identify the most frequent themes related to the paracetamol products being discussed within 18 French forums. The data demonstrated that many individuals use social media to converse regarding paracetamol-containing products, discussing drug efficacy, safety, and more. The Detec't tool captured the information within these discussions and separated it, creating categories to allow for the analysis of particular areas of interest.

The data acquired within this study provide valuable information on web-based discussions involving paracetamol, allowing a better understanding of the needs and concerns of individuals taking or considering taking paracetamol. The demographic analysis within this study appears to concur with previous findings. Of the posts made by people of a known gender, there was more content from women than men, demonstrating a potential bias in the web users. This has been noted within other studies, and it is recognized that women are generally more likely to frequent web forums to discuss health-related information than men [28,29]. The mean age of web users was 39 years, slightly younger than in findings from a US study, which found that the majority of individuals using drug review websites were aged 45 to 64 years [29].

A higher volume of posts was generally seen in colder months than warmer months, possibly linked to cold weather causing an increase or worsening in the prevalence of common illnesses that cause pain or fever [30,31]. The spike in post activity seen within the Paracetamol Only and Paracetamol and Opioids groups between 2009 and 2012 corresponds with the withdrawal of dextropropoxyphene-containing medicines from the market [32,33]. Another spike in post activity seen in 2017 likely corresponds to a change in drug regulations, with all opioid-containing medication requiring a prescription from July 2017 onward [34]. Some of the content within these posts referenced finding alternative products for recreational use, as well as methods for separating paracetamol from opioids in combined medication, supporting the notion that the prescription mandate spike was associated with the discussions around opioid addiction. This is supported by comments discussing addiction and fear of not being able to access the drug, as well as alternative treatments. Given that pain and pyrexia are indications for which paracetamol may be used [27], it was not surprising that pain was the most common medical concept discussed across all posts and pyrexia was the third most

XSL•FO
RenderX

common. Fatigue, the second most common medical concept, was often mentioned in relation to feeling unwell due to illness or sleep disorders caused by pain or fever.

The simultaneous drug consumption analysis found that taking paracetamol in combination with other medications or molecules was common. A study conducted in France found that approximately 23% of individuals were dispensed paracetamol in combination with another agent [2]. Combinations such as including tramadol with paracetamol are common. The different mechanisms of action of the paracetamol and weak opioids are believed to provide improved analgesic efficacy [35]. As an OTC medication, many individuals may not consider consulting a clinician or pharmacist on the risks of taking different OTC drugs and may be unaware that paracetamol is contained in various OTC drugs. This finding highlights the need for information targeting the general public, addressing the possible consequences of taking paracetamol concomitantly with other medications.

In addition to providing general information on paracetamol use, the Detec't tool was also able to identify messages that indicated potential incorrect use of paracetamol, including exceeding the maximum daily recommended dose and dependence. This provides an insight into the views of the individuals who take paracetamol and highlights areas where increased consumer awareness of the dangers of incorrect use could be improved. The number of references to incorrect use suggests that further educational efforts may be required. Many individuals may be unaware of the potential toxicity associated with overdose, and the warning message recently added on the boxes in France should increase consumer awareness of this risk [5,36]. As Doliprane is one of the most popular brands of paracetamol mentioned within the top 10 treatments, it was unsurprising that it was the most popular brand mentioned within the analysis on overdose.

Underreporting of adverse drug events in real-world use has previously been documented, and a 2017 study confirmed this finding, demonstrating that for biologics and drugs with a narrow therapeutic index, approximately 20% to 33% of the minimum number of expected serious events were reported [37]. Social media mining has been used in studies to better understand the consumer mindset and may be able to assist with pharmacovigilance. A study conducted on the public opinion of women with inflammatory bowel disease (IBD) found 1818 posts relating to reproductive concerns for women taking medication for IBD. However, while the women had attributed a risk of reproductive problems to the medication, the condition itself can cause reproductive problems. This valuable information can be used to assist health care professionals to preemptively address these concerns in women taking IBD medication [9]. Within another study investigating methylphenidate for treatment of attention-deficit/hyperactivity disorder, within 3443 social media posts published between 2007 and 2016, 61 adverse events (AEs) were detected along with cases of misuse [13]. A study investigating the use of social media for toxicovigilance purposes found that within 6400 Twitter posts, tweets containing abuse signals were significantly higher for the 3 prescription medications—Adderall (22.6%), quetiapine (5.1%), and oxycodone (12.3%)—compared with metformin (0.3%), a control medication [38]. The WEB-RADAR (Recognizing Adverse Drug Reactions) project [39] investigated the value of data mining Facebook and Twitter for safety signal and AE detection compared with Vigibase [40]. While these social media platforms did not provide valuable pharmacovigilance information, it was suggested that more specialized social media platforms could enrich traditional signal detection, particularly for areas such as pregnancy and drug abuse/misuse [39].

Data mining of social media conversations regarding medication use allows a better understanding of individuals' perceptions and knowledge about medications, often uncovering conversations that are not conducted within the health care setting. These findings may be used to improve health care by preemptively addressing areas of concern and also demonstrate that more easily accessible health care information for the general public would be beneficial. Paracetamol risk, depression, and suicide attempt were the most common discussion topics, suggesting that due to ease of availability, paracetamol consumption in toxic doses for suicidal purposes may be an issue, in line with previously published studies [41,42]. Additionally, discussion topics such as drug use during pregnancy and prescriptions for children indicate that targeting information on the safe and effective use of paracetamol to the public may be beneficial in order to forestall the possible spread of misinformation through online forums. Morbidity and mortality related to self-poisoning can be significantly reduced when a limit is imposed on the sale of paracetamol in a single purchase, suggesting that positive initiatives may assist in protecting the public [43].

A public consultation was launched by ANSM in 2018 with the aim of reducing the number of paracetamol-associated overdoses and medication errors [5]. One of the solutions identified by health authorities was to add a warning message to the product label regarding the risk of overdose with paracetamol, a measure implemented after the end of our study. With the additional information provided by the health authorities, it is hoped that individuals who use paracetamol will have a positive shift in behavior with the increased awareness of the drug's risk [5]. It may be interesting to repeat this study to determine if any changes in how paracetamol is discussed on social media platforms since the introduction of these measures by ANSM can be detected. Such measures should be tailored to the awareness level of the population regarding appropriate use and adverse effects of paracetamol. Whereas a cross-sectional survey conducted in Saudi Arabia showed overall awareness of correct use of paracetamol to be low [44], the advantage of using web-based information collection includes the ability to gather information from individuals who might not otherwise take part in studies [10], as well as the ability to conduct global analyses with real-time collection from a broad sociodemographic range [9,45-47].

Limitations

The statistical algorithms used have been developed for use on large pharmacovigilance databases, opening the possibility of conducting very large pharmacovigilance studies with relative ease. While this method has advantages over traditional

XSL•FO
RenderX

information gathering, there are also limitations, including the fact that this is based on the web user's declaration.

The majority of the posts were made by women, and this potential bias should be considered because women have been shown to display different behavior when searching for information online compared with men and are more likely to consult more sources and value content that is easier to grasp, whereas men tend to prefer a more comprehensive or accurate source [48].

The findings presented are data generated through machine learning predictions. Analyzing the unapproved indications identified within posts containing paracetamol references (eg, the indications) may therefore be viewed as pertaining to paracetamol (ie, linked with paracetamol), or alternatively the data may be unrelated. The style of language used by the web user, including metaphors, slang, or euphemisms, can prove challenging to code as well, although some alternative drug names were included, as shown in Multimedia Appendix 1. Of note, the potential AEs reported are not a reflection of the actual number of AEs; instead, they represent the proportion of posts describing a medical concept that was found to be associated with a potential AE due to drug intake as determined by an algorithm. In addition, some mentions of possible AEs may be confused with drug indications or questions about a possible risk of AE, which would be more likely to be differentiated with additional steps in the analysis, such as human review. Generally, subtleties within the posts may be missed with the use of the machine learning model. Despite these possible sources of bias, the Detec't web-based information collection system used within this study has been previously validated in studies investigating individuals' opinions on drugs via social media and has been found to have good specificity when identifying signals of disproportionate reporting within French medical forums instead of the traditional reporting system [11], including being over 95% accurate in removing posts containing nonadverse drug reactions [16]. However, it should be noted that the results we present are specific for this study and search parameters, so it is not possible to guarantee the algorithm was as successful in removing false positives or negatives. Additionally, the results are specific to France and may not be generalizable to other French-speaking countries or regions due to the cultural sensitivity of some topics.

Conclusion

The use of social media mining with the Detec't tool provided valuable information on the perceptions and understanding of the web users, highlighting areas where providing more information for the general public on paracetamol, as well as other medications, may be of benefit.

Acknowledgments

The study was sponsored by Sanofi. The authors also thank Megan Giliam, MSc, and Emma Watts, PhD, from Lucid Group, Loudwater, UK, for providing medical writing support, which was funded by Sanofi in accordance with Good Publication Practice guidelines. Responsibility for all opinions, conclusions, and interpretation of data lies with the authors. No author was paid for services involved in writing this manuscript.

Conflicts of Interest

SS, AG, PV, and PF are employees of Kap Code. AR was an employee of Kap Code at the time the study was completed. CP and BJ are Sanofi employees.

Multimedia Appendix 1
Forums used for postextraction.
[DOCX File , 22 KB - jmir_v23i7e25049_app1.docx]

Multimedia Appendix 2
List of keywords used for the postextraction categorized by drug composition.
[DOCX File , 24 KB - jmir_v23i7e25049_app2.docx]

Multimedia Appendix 3
Volume of posts containing mention of paracetamol over time for the Paracetamol Only group.
[PNG File , 63 KB - jmir_v23i7e25049_app3.png]

Multimedia Appendix 4
Volume of posts containing mention of paracetamol over time for the Paracetamol and Opioids group.
[PNG File , 62 KB - jmir_v23i7e25049_app4.png]

Multimedia Appendix 5
Volume of posts containing mention of paracetamol over time for the Paracetamol and Others group.
[PNG File , 67 KB - jmir_v23i7e25049_app5.png]

XSL•FO
RenderX

References

1. Bond C, Hannaford P. Issues related to monitoring the safety of over-the-counter (OTC) medicines. Drug Saf 2003;26(15):1065-1074. [doi: 10.2165/00002018-200326150-00001] [Medline: 14640771]
2. Duong M, Gulmez SE, Salvo F, Abouelfath A, Lassalle R, Droz C, et al. Usage patterns of paracetamol in France. Br J Clin Pharmacol 2016 Aug;82(2):498-503 [FREE Full text] [doi: 10.1111/bcp.12957] [Medline: 27059520]
3. Wolf MS, King J, Jacobson K, Di Francesco L, Bailey SC, Mullen R, et al. Risk of unintentional overdose with non-prescription acetaminophen products. J Gen Intern Med 2012 Dec;27(12):1587-1593 [FREE Full text] [doi: 10.1007/s11606-012-2096-3] [Medline: 22638604]
4. King JP, Davis TC, Bailey SC, Jacobson KL, Hedlund LA, Di Francesco L, et al. Developing consumer-centered, nonprescription drug labeling a study in acetaminophen. Am J Prev Med 2011 Jun;40(6):593-598. [doi: 10.1016/j.amepre.2011.02.016] [Medline: 21565649]
5. Paracetamol and risk for the liver: a warning message added to the medicine boxes—press release. Agence Nationale de Sécurité du Médicament et des Produits de Santé. 2019. URL: https://www.ansm.sante.fr/S-informer/Communiques-Communiques-Points-presse/Paracetamol-et-risque-pour-le-foie-un-message-d-alerte-ajoute-sur-les-boites-de-medicament-Communique [accessed 2021-06-15]
6. Perrin A. Social networking usage: 2005-2015. Washington: Pew Internet and American Life Project; 2015 Oct 08. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2015/10/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf [accessed 2021-06-15]
7. Lamas E, Salinas R, Ferrer M, Bousquet C, Vuillaume D. Lay crowd-sourced expertise (LCE) and its influence on the new role of patients: ethical and societal issues. Stud Health Technol Inform 2016;228:80-84. [Medline: 27577346]
8. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013;92(3):426-431 [FREE Full text] [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
9. Keller MS, Mosadeghi S, Cohen ER, Kwan J, Spiegel BMR. Reproductive health and medication concerns for patients with inflammatory bowel disease: thematic and quantitative analysis using social listening. J Med Internet Res 2018 Jun 11;20(6):e206 [FREE Full text] [doi: 10.2196/jmir.9870] [Medline: 29891471]
10. Langer R, Beckman SC. Sensitive research topics: netnography revisited. Qualitative Mrkt Res Int J 2005 Jun;8(2):189-203. [doi: 10.1108/13522750510592454]
11. Kürzinger M, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-based signal detection using medical forums data in France: comparative analysis. J Med Internet Res 2018 Nov 20;20(11):e10466 [FREE Full text] [doi: 10.2196/10466] [Medline: 30459145]
12. Yang M, Kiang M, Shang W. Filtering big data from social media: building an early warning system for adverse drug reactions. J Biomed Inform 2015 Apr;54:230-240 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.011] [Medline: 25688695]
13. Chen X, Faviez C, Schuck S, Lillo-Le-Louët A, Texier N, Dahamna B, et al. Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate. Front Pharmacol 2018;9:541 [FREE Full text] [doi: 10.3389/fphar.2018.00541] [Medline: 29881351]
14. Innovative medicines initiative WEB-RADR workshop report: mobile technologies and social media as new tools in pharmacovigilance. London: European Medicines Agency; 2014 Dec 10. URL: https://web-radr.eu/wp-content/uploads/2015/09/1st-web-radr-workshop-report.pdf [accessed 2021-06-15]
15. FDA Data Mining Council, Duggirala H. Data mining at FDA: white paper. 2018. URL: https://www.fda.gov/media/91848/download [accessed 2021-06-16]
16. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
17. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of cases of noncompliance to drug treatment in patient forum posts: topic model approach. J Med Internet Res 2018 Mar 14;20(3):e85 [FREE Full text] [doi: 10.2196/jmir.9222] [Medline: 29540337]
18. Markov AA. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. In: Howard RA, editor. Dynamic Probabilistic Systems, Volume 1: Markov Chains. Hoboken: John Wiley and Sons; 1971.
19. Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. Ann Math Statist 1966 Dec;37(6):1554-1563. [doi: 10.1214/aoms/1177699147]
20. Cleveland R, Cleveland W, McRae J, Terpenning I. STL: a seasonal-trend decomposition procedure based on loess. J Off Stat 1990;6:3-73 [FREE Full text]
21. Yesha R, Gangopadhyay A. A method for analyzing health behavior in online forums. Proc 6th ACM Conf Bioinformat Computat Biol Health Informat 2015:615-621. [doi: 10.1145/2808719.2812592]
22. Tapi NMD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]

XSL•FO
RenderX

23. Blei DM, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]
24. Blei D, Lafferty J. Correlated topic models. Proc 18th Int Conf Neural Inf Proc Syst 2005:1. [doi: 10.5555/2976248.2976267]
25. Roberts ME, Stewart BM, Tingley D. stm: an R package for structural topic models. J Stat Soft 2019;10(2):1-40 [FREE Full text] [doi: 10.18637/jss.v091.i02]
26. Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Found Trends Inf Retrieval 2009;3(4):333-389 [FREE Full text] [doi: 10.1561/1500000019]
27. [Public medication database: Doliprane 1000 mg, tablet: summary of product characteristics]. French Ministry of Solidarity and Health. 2019. URL: http://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=60234100&typedoc=R [accessed 2021-06-15]
28. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
29. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A study of the demographics of web-based health-related social media users. J Med Internet Res 2015;17(8):e194 [FREE Full text] [doi: 10.2196/jmir.4308] [Medline: 26250986]
30. Linder JE, Kraft DC, Mohamed Y, Lu Z, Heil L, Tollefson S, et al. Human rhinovirus C: age, season, and lower respiratory illness over the past 3 decades. J Allergy Clin Immunol 2013 Jan;131(1):69-77 [FREE Full text] [doi: 10.1016/j.jaci.2012.09.033] [Medline: 23146382]
31. Couch RB, Englund JA, Whimbey E. Respiratory viral infections in immunocompetent and immunocompromised persons. Am J Med 1997 Mar 17;102(3A):2-9 [FREE Full text] [doi: 10.1016/s0002-9343(97)00003-x] [Medline: 10868136]
32. Reset A, Skurtveit S, Furu K, Skovlund E. Effect of the market withdrawal of dextropropoxyphene on use of other prescribed analgesics. Scand J Pain 2018 Oct 25;18(4):667-674. [doi: 10.1515/sjpain-2018-0103] [Medline: 30145581]
33. Press release: European Medicines Agency recommends withdrawal of dextropropoxyphene-containing medicines. European Medicines Agency. 2009. URL: https://www.ema.europa.eu/documents/press-release/european-medicines-agency-recommends-withdrawal-dextropropoxyphene-containing-medicines_en.pdf [accessed 2021-06-15]
34. National Agency for Medicines and Health Products Safety (ANSM). Use and abuse of opioid analgesics: situation report. 2019. URL: https://ansm.sante.fr/actualites/antalgiques-opioides-lansm-publie-un-etat-des-lieux-de-la-consommation-en-france [accessed 2021-06-21]
35. Schug SA. Combination analgesia in 2005—a rational approach: focus on paracetamol-tramadol. Clin Rheumatol 2006;25 Suppl 1:S16-S21. [doi: 10.1007/s10067-006-0202-9] [Medline: 16741784]
36. O'Rourke M, Garland MR, McCormick PA. Ease of access is a principal factor in the frequency of paracetamol overdose. Ir J Med Sci 2002;171(3):148-150. [doi: 10.1007/BF03170503] [Medline: 15736354]
37. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). Expert Opin Drug Saf 2017 Jul;16(7):761-767. [doi: 10.1080/14740338.2017.1323867] [Medline: 28447485]
38. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. Drug Saf 2016 Mar;39(3):231-240 [FREE Full text] [doi: 10.1007/s40264-015-0379-4] [Medline: 26748505]
39. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR. Drug Saf 2019 Dec;42(12):1393-1407 [FREE Full text] [doi: 10.1007/s40264-019-00858-7] [Medline: 31446567]
40. Vigibase: WHO global database of individual case safety reports. Uppsala Monitoring Centre. URL: https://www.who-umc.org/vigibase/vigibase/ [accessed 2021-06-15]
41. Alghamdy MS, Randhawa MA, Al-Wahhas MH, Al-Jumaan MA. Admissions for drug-related problems at the Emergency Department of a University Hospital in the Kingdom of Saudi Arabia. J Family Community Med 2015;22(1):44-48 [FREE Full text] [doi: 10.4103/2230-8229.149590] [Medline: 25657611]
42. Casey D, Geulayov G, Bale E, Brand F, Clements C, Kapur N, et al. Paracetamol self-poisoning: epidemiological study of trends and patient characteristics from the multicentre study of self-harm in England. J Affect Disord 2020 Nov 01;276:699-706. [doi: 10.1016/j.jad.2020.07.091] [Medline: 32871703]
43. Hawton K, Townsend E, Deeks J, Appleby L, Gunnell D, Bennewith O, et al. Effects of legislation restricting pack sizes of paracetamol and salicylate on self poisoning in the United Kingdom: before and after study. BMJ 2001 May 19;322(7296):1203-1207 [FREE Full text] [doi: 10.1136/bmj.322.7296.1203] [Medline: 11358770]
44. Ali SI, Khawaja R, Khan A, Al shamarany A. Perception towards appropriate use and adverse effects of paracetamol among the residents of Al Ahsa. Int J Healthc Sci 2015 Oct;3(2):307-310 [FREE Full text]
45. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]

XSL•FO
RenderX

46. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
47. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. J Big Data 2014;1(1):1-35. [doi: 10.1186/2196-1115-1-2]
48. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. J Assn Inf Sci Tec 2015 Aug 25;68(1):36-47. [doi: 10.1002/asi.23597]

Abbreviations

AE: adverse event
ANSM: Agence Nationale de Sécurité du Médicament et des Produits de Santé (French National Agency for Medicines and Health Products Safety)
EMA: European Medicines Agency
FDA: Food and Drug Administration
HMM: Hidden Markov Model
IBD: inflammatory bowel disease
LOESS: locally estimated scatterplot smoothing
MedDRA: Medical Dictionary for Regulatory Activities
OTC: over-the-counter
PAE: potential adverse event
SmPC: Summary of Product Characteristics
SVM: support vector machine
TF-IDF: term-frequency inverse document frequency
WEB-RADR: Recognizing Adverse Drug Reactions

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by T Freeman, G Martucci, A Rovetta; comments to author 29.01.21; revised version received 24.03.21; accepted 25.04.21; published 12.07.21.

Please cite as:
Schück S, Roustamal A, Gedik A, Voillot P, Foulquié P, Penfornis C, Job B
Assessing Patient Perceptions and Experiences of Paracetamol in France: Infodemiology Study Using Social Media Data Mining
J Med Internet Res 2021;23(7):e25049
URL: https://www.jmir.org/2021/7/e25049
doi:10.2196/25049
PMID:34255645

©Stéphane Schück, Avesta Roustamal, Anaïs Gedik, Paméla Voillot, Pierre Foulquié, Catherine Penfornis, Bernard Job. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Application Programming Interfaces (APIs) to Access Google Data and Gain Insights Into Searches on Birth Control in Louisiana and Mississippi, 2014-2018: Infoveillance Study

Ndola Prata[1], MD, MSc; Karen Weidert[1], MPH; Anne Zepecki[2], BSc; Elina Yon[3], BA; Elizabeth Pleasants[2], MPH; Petrice Sams-Abiodun[4], PhD; Sylvia Guendelman[2], PhD

[1]Bixby Center for Population, Health and Sustainability, School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[2]The Wallace Center for Maternal, Child, and Adolescent Health, School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[3]College of Engineering, University of California, Berkeley, Berkeley, CA, United States

[4]Planned Parenthood Gulf Coast, New Orleans, LA, United States

Corresponding Author:
Ndola Prata, MD, MSc
Bixby Center for Population, Health and Sustainability
School of Public Health
University of California, Berkeley
6142 Berkeley Way West
Berkeley, CA, 94720-6390
United States
Phone: 1 5106426915
Email: ndola@berkeley.edu

Abstract

Background: It is now common to search for health information online. A 2013 Pew Research Center survey found that 77% of online health seekers began their query at a search engine. The widespread use of online health information seeking also applies to women's reproductive health. Despite online interest in birth control, not much is known about related interests and concerns reflected in the search terms in the United States.

Objective: In this study, we identify the top search terms on Google related to birth control in Louisiana and Mississippi and compare those results to the broader United States, examining how Google searches on birth control have evolved over time and identifying regional variation within states.

Methods: We accessed search data on birth control from 2014-2018 from 2 Google application programming interfaces (APIs), Google Trends and Google Health Trends. We selected Google as it is the most commonly used search engine. We focused our analysis on data from 2017 and compared with 2018 data as appropriate. To assess trends, we analyzed data from 2014 through 2018. To compare the relative search frequencies of the top queries across Louisiana, Mississippi, and the United States, we used the Google Health Trends API. Relative search volume by designated marketing area (DMA) gave us the rankings of search volume for each birth control method in each DMA as compared to one another.

Results: Results showed that when people searched for "birth control" in Louisiana and the broader United States, they were searching for information on a diverse spectrum of methods. This differs from Mississippi, where the data indicated people were mainly searching for information related to birth control pills. Across all locations, searches for birth control pills were significantly higher than any other queries related to birth control in the United States, Louisiana, and Mississippi, and this trend remained constant from 2014 to 2018. Regional level analysis showed variations in search traffic for birth control across each state.

Conclusions: The internet is a growing source of health information for many users, including information on birth control. Understanding popular Google search queries on birth control can inform in-person discussions initiated by family planning practitioners and broader birth control messaging campaigns.

International Registered Report Identifier (IRRID): RR2-10.2196/16543

(J Med Internet Res 2021;23(7):e25923) doi:10.2196/25923

XSL•FO
RenderX

KEYWORDS

birth control; search data; Google Trends; infoveillance; infodemiology; Louisiana; Mississippi

Introduction

The use of the internet to source health information has become increasingly popular. It is now common to seek information online prior to, or in some cases, instead of meeting with health professionals [1-3]. A 2013 Pew Research Center survey found that 77% of online health seekers began their query at a search engine [4]. The widespread use of online health information seeking also applies to women's reproductive health [5,6]. One study estimated women's most prevalent health concerns by examining commonly searched health-related keywords in the Google search engine between 2012 and 2013 in the United States. The authors found contraception among the top 25 monthly searched keywords of all categories, with 27,100 related searches estimated each month [5]. These searches were more common in states with the highest number of abortion restrictions. Similarly, a study found that between 2011 and 2015, the number of Google searches in the United States using terms related to self-abortion increased from 119,000 to 700,000 [7], a period of time when states enacted many anti-abortion measures [8]. A qualitative study on the role of social networks in contraceptive decision making revealed that two-thirds of participants used the internet to find information on contraception, noting the privacy of the internet made it an ideal resource [9].

Despite online interest in birth control, not much is known about related interests and concerns reflected in the search terms of the United States population. Online searches represent a novel source of data for understanding contraception needs among internet users. Additionally, the needs for online information and contraceptive resources may be changing as reproductive health policies increasingly restrict access and funding for birth control services and information [7,8,10,11]. In recent years, legislative and regulatory attacks against sexual and reproductive health have reduced the availability of family planning providers. Title X, the only United States federal program dedicated to providing family planning services for low-income people, has narrowed the network of providers by excluding those that offer abortion [7,10].

This study explores online Google search traffic related to birth control in 2 states: Louisiana and Mississippi. We selected these states for our study because they have some of the poorest reproductive health indicators in the country as well as large rural populations and communities of color that worsen their barriers to in-person contraceptive care [12,13]. These 2 states also count on limited data to document population-level contraceptive preferences and concerns [14-16]. Through application of our protocol, we identified the search topics related to birth control that are most often searched on Google in Louisiana and Mississippi and compared those results to the United States as a whole, examining how the searches have changed over time. Furthermore, we examined variations in birth control searches across geographic areas within each state to explore how search interest for each contraceptive method varies between different geographical areas. We hypothesized that understanding the information people are seeking online can help strengthen the capacity of key organizations and reproductive health leaders to deliver appropriate, acceptable, client-centered, and high-quality contraceptive services. Data on the contraceptive methods most searched by geolocation can inform information, education, and communications campaigns and ensure potential users are getting clear information on various methods.

Methods

We accessed search data on birth control from 2014 to 2018 from 2 Google application programming interfaces (APIs), Google Trends and Google Health Trends. We selected Google as it is the most commonly used search engine [17]. We focused our analysis on data from 2017, and comparisons with 2018 data were made as appropriate. To assess trends, we analyzed data from 2014 through 2018.

Figure 1 illustrates the steps in retrieving data from the Google APIs. We followed a methodological framework to retrieve Google data from multiple APIs [18]. Google Trends API allowed us to retrieve search queries linked to the initial search term "birth control" for a given location in 2017.

Figure 1. Steps in retrieving data from custom Google application programming interfaces (APIs).



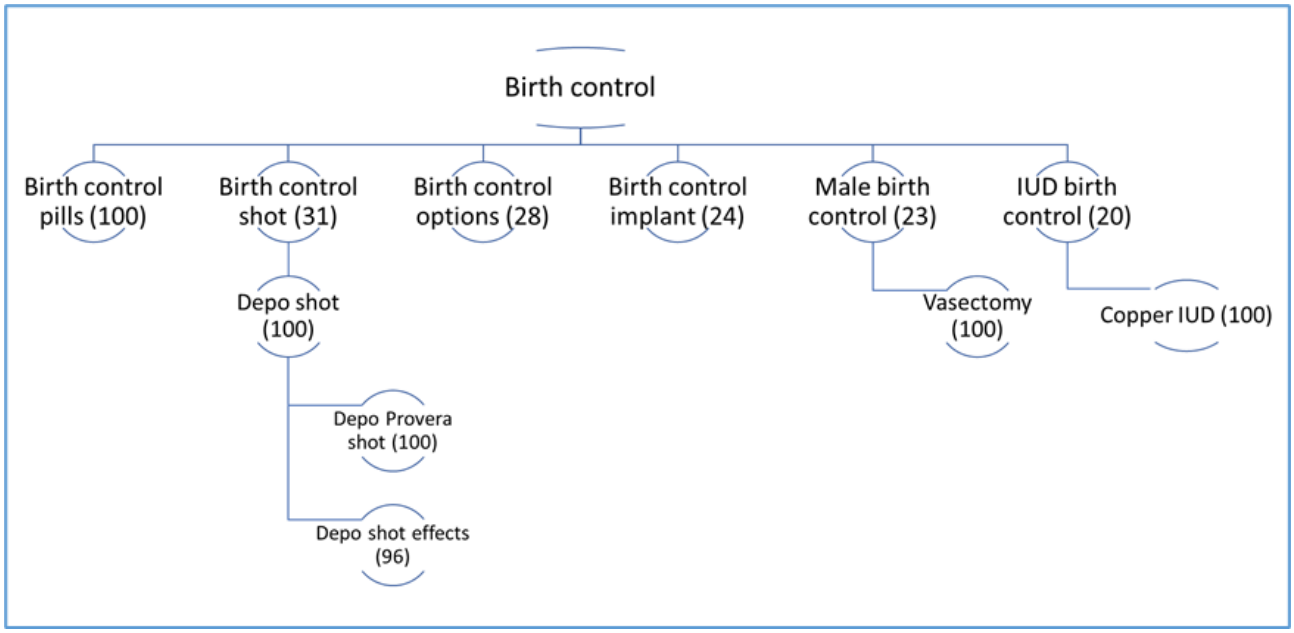
We arrived at "birth control" as our preferred search term after implementing a simulation methodology [18] that showed very strong association with birth control methods, compared to weak or no association between "contraception" or "family planning" with birth control methods. The Google Trends API returns the top queries a user searches along with a relative search index ranging from 0 to 100 (with 100 denoting the query with the strongest association to the initial search term and 0 denoting the weakest association). The Google Health Trends API normalizes search data to make comparisons between queries easier. Search results are normalized to the time and location of a query, by dividing each data point by the total searches of the geography and time range it represents to compare relative volume. The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics. Google Trends does filter out some types of searches, such as

XSL•FO
RenderX

duplicate searches done by the same person and those made by very few people; Google Trends only shows data for popular terms, so search terms with low volume appear as "0" [19]. Top queries are not displayed when the search volume for a given query falls below what Google considers the threshold of traffic volume (search terms are assigned a value of 0); the search volume often does not meet this threshold in less densely populated areas. We created a master list that included all of the top queries from the United States, Louisiana, and Mississippi and the most strongly associated follow-up terms that derived from the top queries. Directed graphs mapped the initial search term to its top related queries in order of the strength of their association, a number presented in parenthesis in each node (Figure 2, Figure 3, and Figure 4).

To compare the relative search frequencies queries across Louisiana, Mississippi, and the United States, we used the Google Health Trends API. The Google Health Trends API did not give results if there are not enough searches above the privacy threshold matching our parameters. We display the relative search volume of the top queries in each location and year in a bar graph (Figure 5). To establish significant differences in relative search volume searches in years prior to 2017, we estimated 95% confidence intervals of the point estimate of each term. However, these significant differences are reported in the results but not in Figure 5.

Figure 2. Directed graph of the top birth control–related search queries in the United States in 2017, with the number value representing the association between the search query and the top topic (birth control), with 100 reflecting a high association.



XSL•FO
RenderX

Figure 3. Directed graph of the top birth control–related search queries in Louisiana in 2017, with the number value representing the association between the search query and the top topic (birth control), with 100 reflecting a high association.

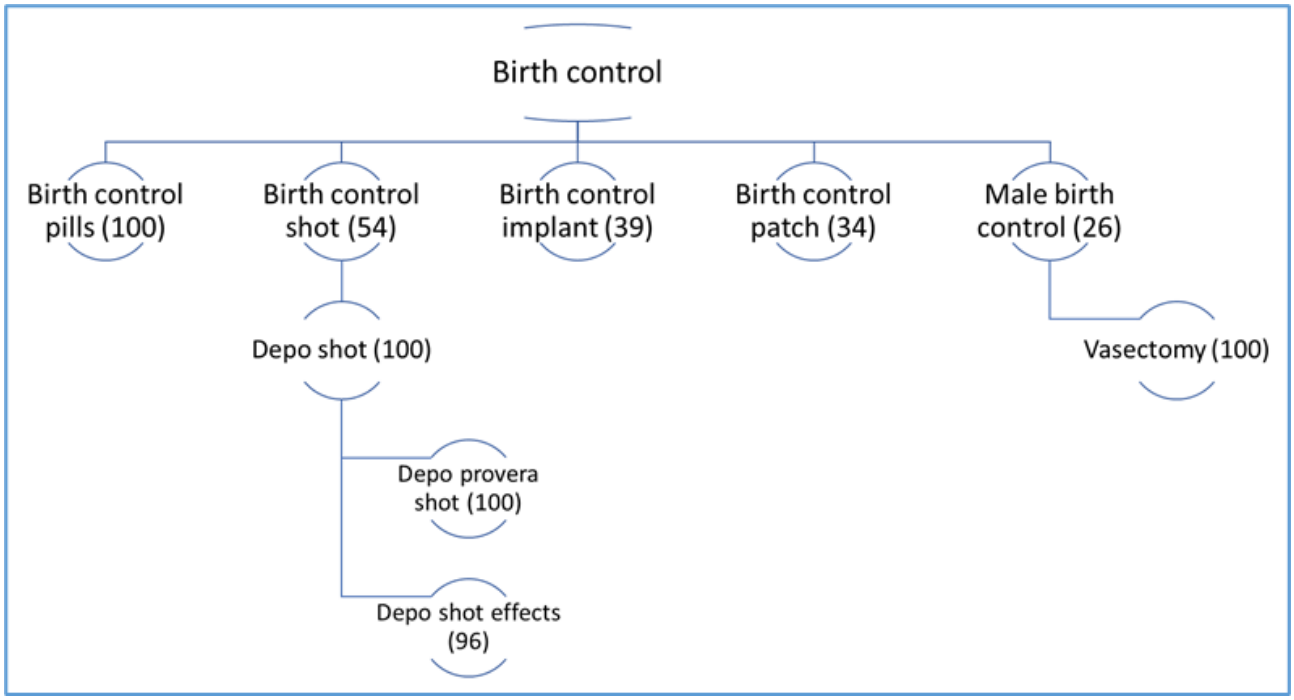
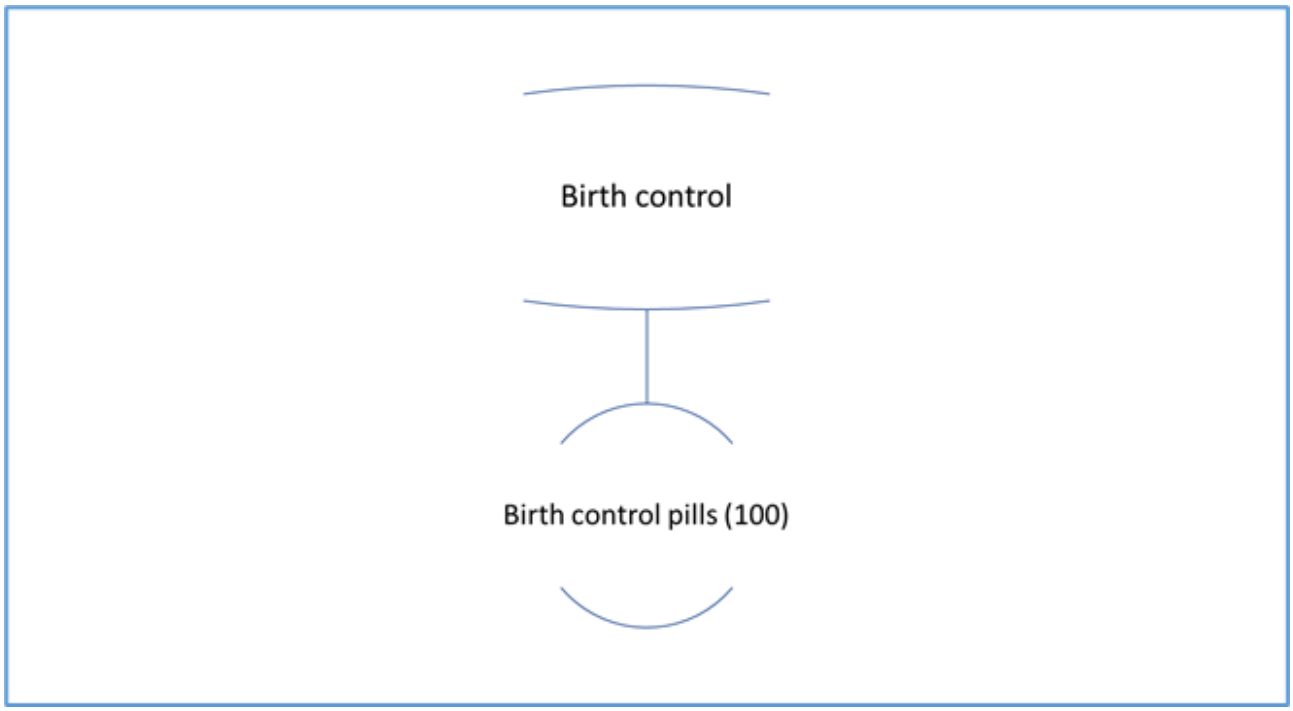
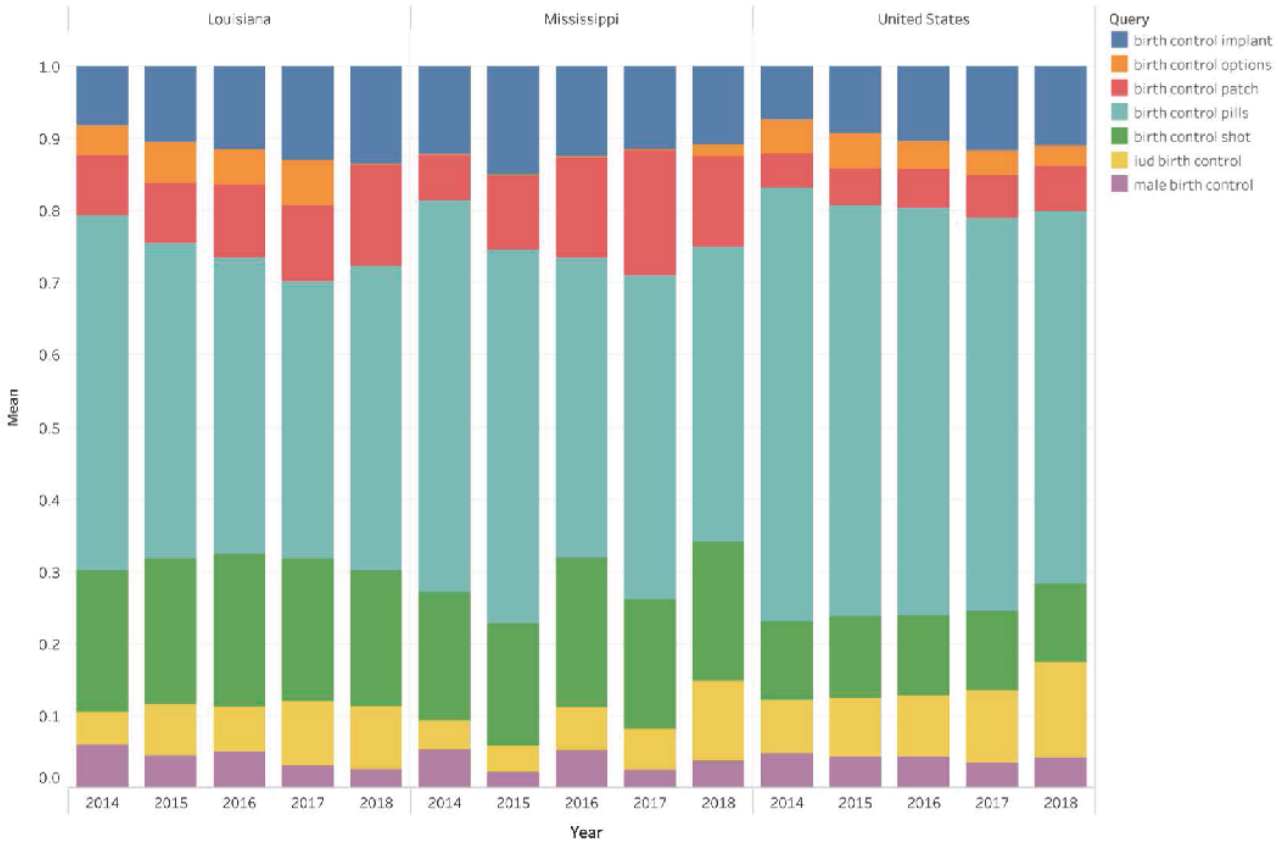


Figure 4. Directive graph of the top birth control–related search queries in in Mississippi in 2017, with the number value representing the association between the search query and the top topic (birth control), with 100 reflecting a high association.



XSL•FO
RenderX

Figure 5. Trends in relative search volumes for top search queries related to birth control in the United States, Louisiana, and Mississippi, 2014-2018; the relative search volume is the proportion of a specific search term among all queries associated with birth control in a given location and year.



To examine regional differences within each state, we focused on designated market areas (DMAs), which are the smallest geolocations for which Google search data are reported. DMAs, which we acquired from Nielsen Company 2017 data, are the geographic areas in the United States in which local television viewing is measured by Nielsen, an American information, data, and measurement firm [20]. DMA data are essential for any marketer, researcher, or organization seeking to utilize standardized geographic areas [21]. Both Louisiana and Mississippi are divided into 7 DMAs; however, for this study, only 5 DMAs in each state met the Google Health Trends API confidentiality threshold for extracting data. Relative search volume by DMA gave us the rankings of search volume for each birth control method in each DMA as compared to one another when sufficient volume of searches was available in the DMA (Figure 6 and Figure 7). These data allowed us to explore how search interest for each contraceptive method varies between different geographical areas in each of the 2 states.

XSL•FO
RenderX

Figure 6. Relative search volume for birth control search queries in Louisiana and in 5 designated marketing areas (DMAs) in Louisiana in 2017. In Louisiana, of the 7 DMAs in total, Alexandria and Lake Charles did not meet the threshold for data extraction and were excluded.

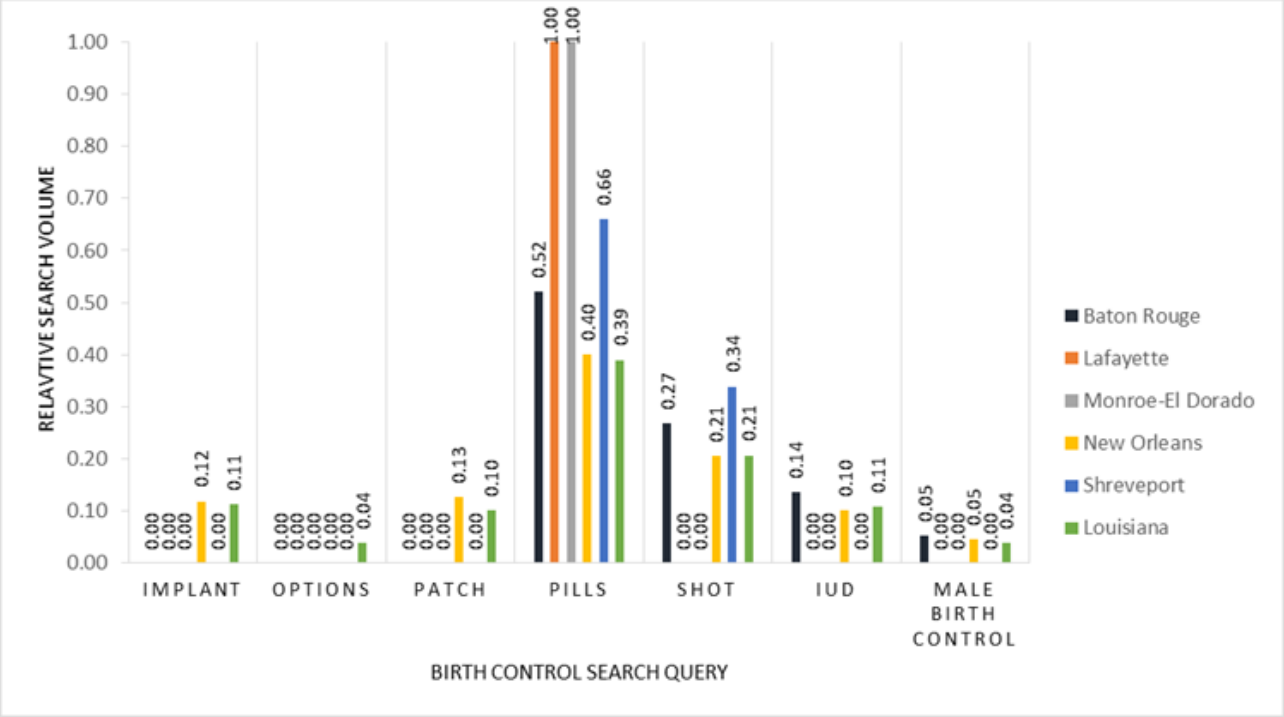
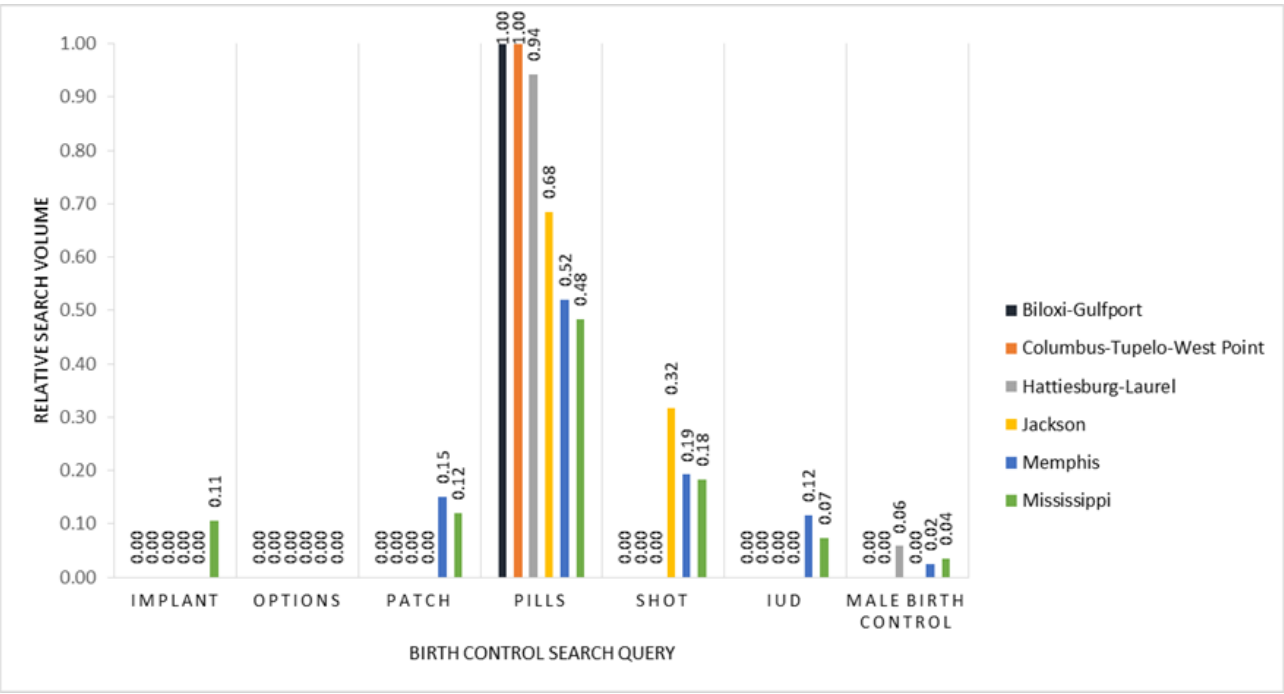


Figure 7. Relative search volume for birth control search queries in Mississippi and in 5 designated marketing areas (DMAs) in Mississippi in 2017. In Mississippi, of the 7 DMAs in total, Greenwood-Greenville and Meridian did not meet the threshold for data extraction and were excluded.



Some of the data presented in this paper were used to test our methodology for retrieving Google data from APIs as published in Journal of Medical Internet Research (JMIR) research protocols (RR1-10.2196/16543) [18]. This paper presents the final analysis and results, although some of the figures might seem similar to those already published alongside the research protocol. The protocol was also applied to abortion searches on Google in Louisiana and Mississippi in a PLOS One publication by Guendeleman et al [22].

Results

Tracking Search Queries on Birth Control

In Figure 2, the directed graph of top queries in the United States as described in the Methods section shows that in 2017, the most popular search query for the term for “birth control” in the United States was birth control pills. The value of 100 for birth control pills denotes that they were most strongly associated with the search term “birth control.” Following birth

XSL•FO
RenderX

control pills, other popular search queries were for the birth control shot (injectable contraceptive), the implant, male birth control (referring only to vasectomy), and IUD (intrauterine device; referring to copper IUD).

Figure 3, the directed graph of top queries in Louisiana, shows that Louisiana followed a similar pattern to the United States, except that IUD was not a popular search query associated with birth control in 2017. Meanwhile, the birth control patch was among the top queries associated with birth control in Louisiana, but not found in the United States population as a whole. Figure 4, the directed graph of top queries in Mississippi, shows that in contrast to the wide range of contraceptive methods frequently searched for in the United States and Louisiana, the top searches in Mississippi were circumscribed to birth control pills. In 2018, the top queries for birth control remained the same in Louisiana and Mississippi (data not shown). However, in the United States, while the IUD was among the top queries in 2017, it disappeared from the top query list in 2018 (data not shown). Even though the IUD continued to be searched, it was not part of the top query list associated with birth control.

Trends in Birth Control Search Traffic Between 2014 and 2018

Unlike search queries on birth control where only those meeting a threshold appeared in the graphs, we employed the Google Health Trends API to include all birth control–related top queries (first row of Figure 2) in the United States to enable comparisons in relative search volume. Relative search volume gives us the proportion of specific search terms among all queries associated with birth control in a given location and year. Figure 5 shows the trends in relative search volume for the birth control top queries during 2014-2018. In the United States, Louisiana, and Mississippi, birth control pills remained the most searched query from 2014 to 2018. While similar relative search volumes through the observed period were seen in the United States and Mississippi, in Louisiana, birth control pills search volume declined from 2014 to 2016 and was similar in 2017 and 2018.

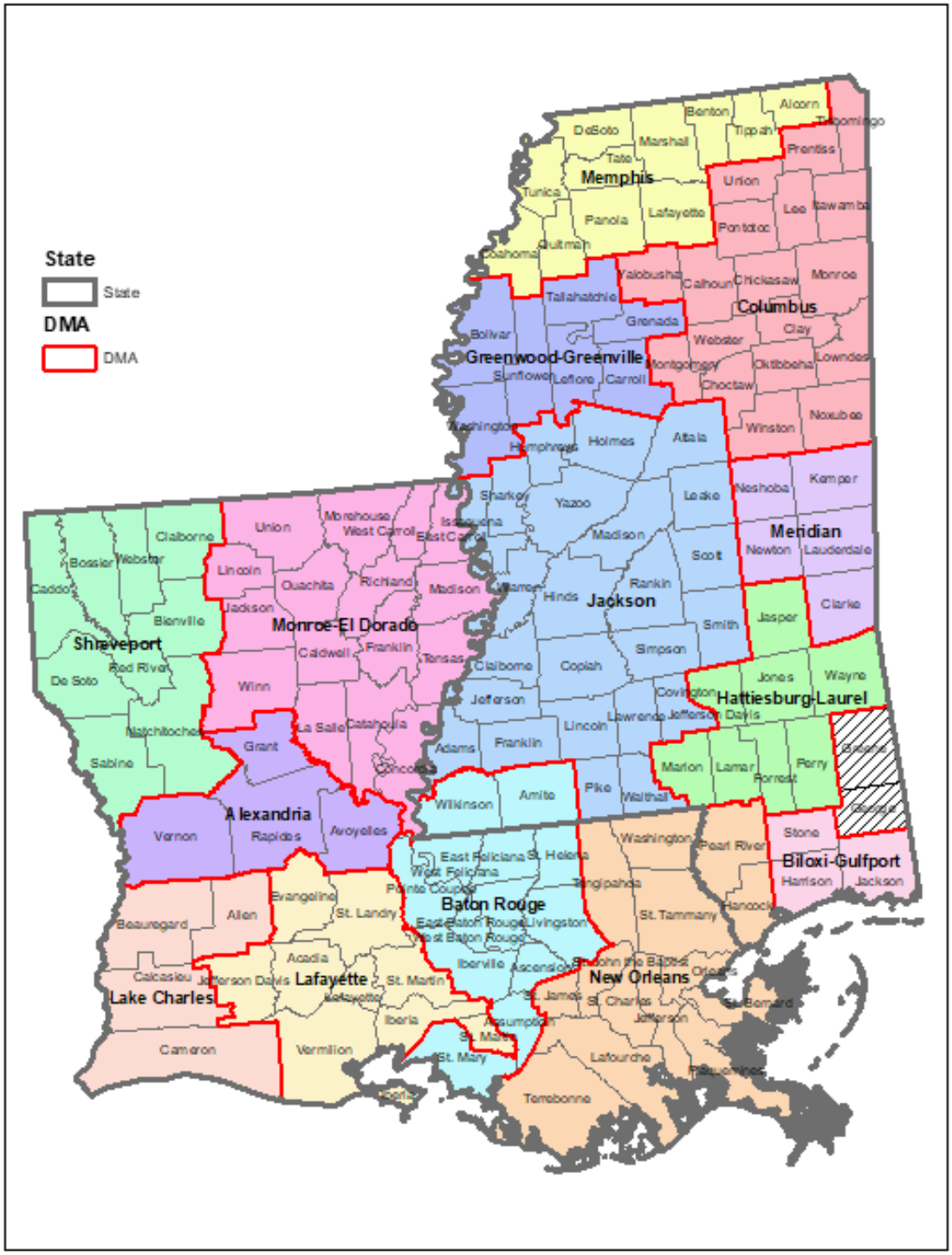
In the United States, Louisiana, and Mississippi, the birth control shot appeared as the second most searched query but not significantly different throughout the period in any of the geolocations (Figure 5). Implants were the third most searched method in Louisiana and Mississippi, while in the United States, it was the IUD. Relative search volume for implants increased over time in Louisiana and the United States, while it remained stable in Mississippi. Increases in relative search volumes for IUDs and the patch were observed in all 3 geolocations in the last 2 years of the studied period (2017-2018).

State-Level Geographic Differences in Birth Control Search Traffic

The DMA-level analysis showed variations in search traffic for birth control methods across each state. Figure 8 breaks down Louisiana and Mississippi by county and DMA. In Louisiana, searches for birth control pills were relatively higher in Lafayette and Monroe El Dorado compared to other DMAs, and birth control pills were the only query related to birth control (Figure 6). Regarding searches on birth control shot, the second most common query, Shreveport emerged as the DMA with the relative highest volume followed by Baton Rouge (Figure 6). The New Orleans DMA had the most diverse queries, similar to the broader Louisiana search pattern. In Mississippi, the searches for the pill were much higher in Biloxi-Gulfport and Columbus; Jackson and Memphis DMAs had higher relative search volumes for the birth control shot (Figure 7). Memphis DMA had the most diverse queries, whereas Biloxi-Gulfport and Columbus only had queries on birth control pills. Regarding searches for IUDs, Memphis represented the majority of searches in the state for this method.

XSL•FO
RenderX

Figure 8. Breakdown of Louisiana and Mississippi counties by designated marketing area (DMA); all 7 DMAs in Mississippi represent aggregated counties or zip codes in the state. In Mississippi, 6 counties (George, Green, Pearl River, Hancock, Amite, and Wilkinson) are located in DMAs that are primarily made up of zip codes in other states and were excluded.



Discussion

Principal Findings

The internet is growing as a platform to source health information for many users, including birth control, as found in this study. While we do not know why people are searching for birth control information on Google, we surmise that it follows a similar pattern to other online health-seeking behavior, with the search engine providing a starting point for answering basic questions, understanding options available, and seeking health services [1,2,23]. Limited access to information and services through other channels may also force individuals to rely on the internet more frequently. Currently, neither Louisiana nor Mississippi requires coverage of prescription contraception

XSL•FO
RenderX

or no-cost contraception coverage [11]. At the same time, while sex education is mandated in Mississippi, the curriculum does not have to be evidence-based or provide information beyond abstinence. In Louisiana, no sex education is mandated, and if provided, it must emphasize abstinence [24]. Lack of access to information and resources has been linked to internet use for information and services. The use of the web may become more important in contraceptive decision making, providing opportunities for individuals to connect with trusted and accurate online resources [9,25]. Experts acknowledge that better-informed choices lead to better reproductive health outcomes [26]. A key factor of preventative health practice is access to information regarding risks to health and promotional measures for improving health status. Meanwhile, the digital age has changed the landscape of health information seeking [27]. A qualitative study to evaluate the process of online health information searching found that participants viewed the internet as a valuable tool for finding health information in order to support their existing health care resources [23]. The authors emphasized knowing what health information people are searching for online can help address knowledge gaps between providers and patients.

We found birth control pills were the most common search query related to birth control across all 3 geolocations. Though we cannot compare searches on birth control to actual use of methods, these findings are supported by a 2017 Guttmacher Institute report assessing state-level estimates of contraceptive use that found that the birth control pill was one of the most commonly used primary methods of contraception among women aged 18-49 years at risk of unintended pregnancy in the United States, ranging from 11% in Alaska to 27% in Massachusetts [28]. In Louisiana and Mississippi, 17.9% and 14.2%, respectively, of women at risk of unintended pregnancy were using birth control pills as their primary method of contraception.

Birth control pills also remained the most popular query throughout the data looking at trends from 2014-2018. Though there have been significant increases in use of long-acting reversible contraception (LARC) in the last decade in the United States [29] and our study showed noticeable increases in searches conducted for LARC over time (2014-2018), actual LARC use in Louisiana and Mississippi is lower than most other states [28]. Louisiana had the lowest use of the IUD at 4.5% of women at risk of unintended pregnancy in the 2017 Guttmacher Institute report. At 7.1% IUD use among women at risk of unintended pregnancy, Mississippi was also on the low end among states; neither Louisiana nor Mississippi had reliable data on implant use. A 2016 study among women of reproductive age in the United States found that 33% of surveyed women were either unaware of LARC (including both IUDs and implants) or had misperceptions about their effectiveness or safety [30]. Another study found there was a sharp increase (21.6%) in the rate of women who chose LARC methods in the 30 days after the 2016 United States presidential election as compared to the 30 days before [31], which corresponded with a spike in social media calling for women to get LARC methods after the election of President Trump given concerns about the future of the Affordable Care Act and contraception coverage [32]. This suggests that Google search patterns may be influenced by mainstream and social media. Additionally, changes to contraceptive care laws and funding, trending news stories, and political events are other factors that might affect online search behaviors.

While we found wide regional variations in search traffic for birth control methods by DMA for both Louisiana and Mississippi, more data collection efforts are needed to identify the factors that are driving differences in search behaviors related to birth control in these localities. However, there were notable differences between DMAs that include larger cities versus the DMAs that were more rural. The New Orleans DMA in Louisiana and the Memphis DMA both contain large cities and also had the most diverse search queries among all DMAs. Though we cannot draw conclusions from these findings, they are important to explore further given both health disparities in rural women [33] and the digital gap between rural and nonrural communities in the United States [34]. Further examination of the regional differences in birth control searches can inform providers and advocates of the birth control interests of the populations in particular geographies. Internet search histories by geography can also be used in combination with service availability and service utilization data to provide a better picture of where the gaps in both services and need for more knowledge exist, including locations to obtain counseling and services.

Limitations

There are several limitations to account for when interpreting the data and methodology. This study focused on the United States, specifically 2 states — Louisiana and Mississippi. Therefore, applicability outside of the United States is limited. Additionally, there is variability in online access and internet usage across different demographic, socioeconomic, and geographic subpopulations within the United States, such that certain groups of users might be over- or underrepresented among internet search data. We do not know the reasons that prompt individuals to search for contraception information, nor the demographic characteristics of the individuals searching for the topics assessed in this analysis. However, previous research has indicated that female-identifying internet users are more likely to seek health information online [4,35]. We also cannot be certain that searching for a topic online reflects intention to use or current use of contraception or how people perceive or use this information. The Google Trends API used for identifying top queries only shows the queries highly associated with the search term “birth control.” Thus, queries that have weaker associations with birth control are not reported. The Google Health Trends API also does not report relative search volume below a certain threshold (unknown to us). Another limitation is related to interpretation of the DMA data. A DMA region is comprised of counties that form an exclusive geographic area in which the home market television stations hold a dominance of total hours viewed, but these counties are not necessarily located in the same state. For example, the Memphis DMA consists of counties in Mississippi, Arkansas, and Tennessee. In Mississippi, 6 counties (George, Green, Pearl River, Hancock, Amite, and Wilkinson) are located in DMAs that are primarily made up of zip codes in other states and therefore were not included in the analysis. In the Louisiana

XSL•FO
RenderX

DMAs, 4 counties were included that are not part of Louisiana. Wilkinson and Amite are located in the Baton Rouge DMA. Pearl River and Hancock are located in the New Orleans DMA. Despite these limitations, the analysis of Google search traffic data presents important insights into the most popular searches related to birth control, information that can be useful for health providers and program implementers in the United States, including those dedicated to providing accurate information about birth control methods.

Conclusions

The convenience, accessibility, and availability of information on the internet have resulted in increasing numbers of individuals searching for health information online, and as shown in our data from the United States, Louisiana, and Mississippi, this includes searching for information on birth control. Understanding popular Google search queries on birth control can inform in-person discussions initiated by family planning practitioners and inform broader birth control messaging campaigns. Further research is needed to understand the quality and quantity of birth control information on the internet and ensure those searching are getting comprehensive and accurate information on birth control.

Acknowledgments

This work was primarily supported by the Lucille and David Packard Foundation (grant number 2017-66708) and partially supported by the Bixby and Wallace Centers in the School of Public Health at the University of California, Berkeley.

Conflicts of Interest

None declared.

References

1. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
2. Gualtieri LN. The doctor as the second opinion and the internet as the first. 2009 Presented at: CHI '09 Extended Abstracts on Human Factors in Computing Systems; April 4-9, 2009; Boston, MA p. 2489-2498. [doi: 10.1145/1520340.1520352]
3. Fox N, Ward K, O'Rourke AJ. The 'expert patient': empowerment or medical dominance? The case of weight loss, pharmaceutical drugs and the Internet. Soc Sci Med 2005 Mar;60(6):1299-1309. [doi: 10.1016/j.socscimed.2004.07.005] [Medline: 15626525]
4. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-06-20]
5. Baazeem M, Abenhaim H. Google and Women's Health-Related Issues: What Does the Search Engine Data Reveal? Online J Public Health Inform 2014 Oct 16;6(2):e187 [FREE Full text] [doi: 10.5210/ojphi.v6i2.5470] [Medline: 25422723]
6. Vogel L. Canadian women opting for less effective birth control. CMAJ 2017 Jul 10;189(27):E921-E922 [FREE Full text] [doi: 10.1503/cmaj.1095446] [Medline: 28694316]
7. State Family Planning Funding Restrictions. Guttmacher Institute. 2020. URL: https://www.guttmacher.org/state-policy/explore/state-family-planning-funding-restrictions [accessed 2021-06-20]
8. Sonfield A. Seeing the whole pattern: coordinated federal attacks on birth control coverage and access. Guttmacher Institute. 2020 Jun 26. URL: https://www.guttmacher.org/article/2020/06/seeing-whole-pattern-coordinated-federal-attacks-birth-control-coverage-and-access [accessed 2021-06-20]
9. Yee L, Simon M. The role of the social network in contraceptive decision-making among young, African American and Latina women. J Adolesc Health 2010 Oct;47(4):374-380 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.03.014] [Medline: 20864007]
10. Dawson R. Trump administration's domestic gag rule has slashed the Title X network's capacity by half. Guttmacher Institute. 2020 Feb 05. URL: https://www.guttmacher.org/article/2020/02/trump-administrations-domestic-gag-rule-has-slashed-title-x-networks-capacity-half [accessed 2021-06-20]
11. State Requirements for Insurance Coverage of Contraceptives. Kaiser Family Foundation. 2020 Aug 01. URL: https://www.kff.org/other/state-indicator/state-requirements-for-insurance-coverage-of-contraceptives/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D [accessed 2021-06-20]
12. Bennett T. Reproductive Health Care in the Rural United States. JAMA 2002 Jan 02;287(1):112. [doi: 10.1001/jama.287.1.112-jms0102-6-1]
13. Howell M, Starrs AM. For women of color, access to vital health services is threatened. Guttmacher Institute. 2017 Jul 27. URL: https://www.guttmacher.org/article/2017/07/women-color-access-vital-health-services-threatened [accessed 2021-06-20]
14. Annual Report 2018: A call to action for individuals and their communities. America's Health Rankings: United Health Foundation. 2018. URL: https://assets.americashealthrankings.org/app/uploads/ahrannual-2018.pdf [accessed 2021-06-20]

XSL•FO
RenderX

15. State Facts About Unintended Pregnancy: Louisiana. Guttmacher Institute. 2014. URL: https://www.guttmacher.org/sites/default/files/factsheet/la_5.pdf [accessed 2021-06-20]
16. Preventing Unintended Pregnancy in Mississippi. Center for Mississippi Health Policy. 2018 May 23. URL: https://mshealthpolicy.com/preventing-unintended-pregnancies/ [accessed 2021-06-20]
17. Share of search queries handled by leading US search engine providers as of January 2020. Statista. URL: https://www.statista.com/statistics/267161/market-share-of-search-engines-in-the-united-states/ [accessed 2021-06-20]
18. Zepecki A, Guendelman S, DeNero J, Prata N. Using Application Programming Interfaces to Access Google Data for Health Research: Protocol for a Methodological Framework. JMIR Res Protoc 2020 Jul 06;9(7):e16543 [FREE Full text] [doi: 10.2196/16543] [Medline: 32442159]
19. FAQ about Google Trends data. Google. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2021-06-20]
20. About Nielsen. The Nielsen Company. URL: https://www.nielsen.com/us/en/about-us/ [accessed 2021-06-20]
21. DMA Regions. The Nielsen Company. URL: https://www.nielsen.com/us/en/contact-us/intl-campaigns/dma-maps/ [accessed 2021-06-20]
22. Guendelman S, Yon E, Pleasants E, Hubbard A, Prata N. Shining the light on abortion: Drivers of online abortion searches across the United States in 2018. PLoS One 2020 May 21;15(5):e0231672 [FREE Full text] [doi: 10.1371/journal.pone.0231672] [Medline: 32437369]
23. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. J Med Internet Res 2014 Oct 07;16(10):e224 [FREE Full text] [doi: 10.2196/jmir.3341] [Medline: 25348028]
24. Sex and HIV Education. Guttmacher Institute. URL: https://www.guttmacher.org/state-policy/explore/sex-and-hiv-education [accessed 2021-06-20]
25. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. Public Health Rep 2019 Sep 12;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]
26. Dehlendorf C, Tharayil M, Anderson N, Gbenedio K, Wittman A, Steinauer J. Counseling about IUDs: a mixed-methods analysis. Perspect Sex Reprod Health 2014 Sep 13;46(3):133-140 [FREE Full text] [doi: 10.1363/46e0814] [Medline: 24628710]
27. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
28. Douglas-Hall A, Kost K, Kavanaugh ML. State-Level Estimates of Contraceptive Use in the United States. Guttmacher Institute. 2018 Dec. URL: https://www.guttmacher.org/report/state-level-estimates-contraceptive-use-us-2017 [accessed 2021-06-20]
29. Kavanaugh M, Jerman J, Finer LB. Changes in Use of Long-Acting Reversible Contraceptive Methods Among U.S. Women, 2009-2012. Obstet Gynecol 2015 Nov;126(5):917-927 [FREE Full text] [doi: 10.1097/AOG.0000000000001094] [Medline: 26444110]
30. Shartzer A, Courtot B, McMorrow S, Benatar S, Kenney GM. Knowledge Gaps and Misinformation about Birth Control Methods Persist in 2016. Urban Institute. 2016. URL: https://www.urban.org/research/publication/knowledge-gaps-and-misinformation-about-birth-control-methods-persist-2016 [accessed 2021-06-20]
31. Pace LE, Dusetzina SB, Murray Horwitz ME, Keating NL. Utilization of Long-Acting Reversible Contraceptives in the United States After vs Before the 2016 US Presidential Election. JAMA Intern Med 2019 Mar 01;179(3):444-446 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7111] [Medline: 30715075]
32. Judge CP, Borrero S. Contraceptive decision making after the 2016 US presidential election. Am J Obstet Gynecol 2017 Jul;217(1):89-90 [FREE Full text] [doi: 10.1016/j.ajog.2017.03.025] [Medline: 28366732]
33. Committee on Health Care for Underserved Women. ACOG Committee Opinion No. 586: Health disparities in rural women. Obstet Gynecol 2014 Feb;123(2 Pt 1):384-388. [doi: 10.1097/01.AOG.0000443278.06393.d6] [Medline: 24451676]
34. Perrin A. Digital gap between rural and nonrural America persists. Pew Research Center. 2019 May 31. URL: https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/ [accessed 2021-06-20]
35. Bach RL, Wenz A. Studying health-related internet and mobile device use using web logs and smartphone records. PLoS One 2020 Jun 12;15(6):e0234663 [FREE Full text] [doi: 10.1371/journal.pone.0234663] [Medline: 32530937]

Abbreviations

API: application programming interface
DMA: designated market area
IUD: intrauterine device
JMIR: Journal of Medical Internet Research
LARC: long-acting reversible contraception

XSL•FO
RenderX

Edited by R Kukafka; submitted 26.11.20; peer-reviewed by L Kruesi, R Hauser; comments to author 05.01.21; revised version received 25.01.21; accepted 04.05.21; published 12.07.21.

Please cite as:
Prata N, Weidert K, Zepecki A, Yon E, Pleasants E, Sams-Abiodun P, Guendelman S
Using Application Programming Interfaces (APIs) to Access Google Data and Gain Insights Into Searches on Birth Control in Louisiana and Mississippi, 2014-2018: Infoveillance Study
J Med Internet Res 2021;23(7):e25923
URL: https://www.jmir.org/2021/7/e25923
doi:10.2196/25923
PMID:34255662

©Ndola Prata, Karen Weidert, Anne Zepecki, Elina Yon, Elizabeth Pleasants, Petrice Sams-Abiodun, Sylvia Guendelman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Influence of Online Health Information Seeking Before a Consultation on Anxiety, Satisfaction, and Information Recall, Mediated by Patient Participation: Field Study

Melanie de Looper[1], MSc; Julia C M van Weert[1], Prof Dr; Barbara C Schouten[1], PhD; Sifra Bolle[2], PhD; Eric H J Belgers[3], Dr med; Eric H Eddes[4], Dr med; Ellen M A Smets[5], Prof Dr

[1]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[2]Minddistrict, Amsterdam, Netherlands

[3]Zuyderland Medical Center, Sittard-Geleen, Netherlands

[4]Deparment of Surgery, Deventer Hospital, Deventer, Netherlands

[5]Department of Medical Psychology, Amsterdam UMC, Amsterdam Public Health, Amsterdam, Netherlands

Corresponding Author:
Melanie de Looper, MSc
Amsterdam School of Communication Research
University of Amsterdam
Nieuwe Achtergracht 166
Amsterdam, 1018 WV
Netherlands
Phone: 31 630274301
Email: m.delooper@uva.nl

Abstract

Background: Today, many cancer patients engage in online health information seeking (OHIS). However, little is known about how patients differ in their OHIS levels. In addition, OHIS might influence patient participation during a consultation with a physician, which might mediate the effects on patient outcomes.

Objective: The aim of this study is twofold: first, to provide insight into which personal characteristics and psychosocial factors affect patients' OHIS levels and, second, to test the hypothesis that the effects of OHIS on patient outcomes are mediated by patient participation during the consultation.

Methods: Patient participation was operationalized in terms of patients' absolute word count; the relative contribution of the patient, compared with the health care provider; and the number of questions and assertions expressed during the consultation. The patient outcomes measured were anxiety after the consultation, satisfaction with the consultation, and information recall. Participants in this study were patients recently diagnosed with colorectal cancer recruited from 6 hospitals in the Netherlands (n=90). Data were collected using questionnaires and audio-recorded consultations of patients with health care providers before their surgery.

Results: The results showed that younger patients, higher educated patients, patients with a monitoring coping style, and patients who experienced more cancer-related stress engaged more in OHIS. In turn, OHIS was related to patient participation in terms of the patient's absolute word count but not to the relative contribution to the consultation or expressing questions and assertions. We did not find a relation between OHIS and anxiety and OHIS and recall mediated by patient participation. However, we found that patients' absolute word count significantly mediated the positive association between OHIS and patients' satisfaction with the consultation.

Conclusions: Results indicate positive implications of OHIS for patients' care experience and, therefore, the importance of helping patients engage in OHIS. However, the results also suggest that OHIS is only successful in increasing a single aspect of patient participation, which might explain the absence of relations with anxiety and recall. The results suggest that more beneficial effects on patient outcomes may be achieved when health care providers support patients in OHIS.

(J Med Internet Res 2021;23(7):e23670) doi:10.2196/23670

XSL•FO
RenderX

KEYWORDS

online health seeking; patients; aging; patient participation; memory; anxiety; patient reported outcomes; consultation; health communication; cancer

Introduction

Background

Today, the internet hosts a growing body of easily accessible cancer-related information [1]. In line with this, cancer patients increasingly engage in online health information seeking (OHIS) [2,3] about their illness and treatment [4]. OHIS about one's health or medical condition can contribute to feeling informed, which has been positively associated with patient outcomes [5]. For instance, better informed patients score higher on affective outcomes, for example, they are more satisfied with their treatment [6-9] and feel less anxious [5,10,11]. Moreover, OHIS can positively influence cognitive outcomes, such as better information recall [12,13].

Although patients generally seek web-based health information [14-16], it can be argued that the extent to which they engage in OHIS is associated with individual differences based on demographics or psychosocial characteristics [17,18]. For instance, experiencing feelings of anxiety or stress regarding a medical diagnosis can result in more information needs [19] and information seeking to cope with them [20].

Previous research did not look at the whole path from individual differences to OHIS and, in turn, patient outcomes but mainly focused on either predictors of OHIS in terms of demographics and psychosocial factors [21-24] or outcomes of OHIS [5,25-27]. More specifically, research that looked into the effects of OHIS did not take into account what happens between OHIS and patient outcomes in terms of consultations with health care providers [5,25]. This is a noteworthy omission because patients often engage in OHIS in preparation for consultations [15,16,28], which can result in a better informed and more empowered patient who feels comfortable in taking on an active role in consultations with health care providers [9,27,29]. In turn, this may lead to more active patient participation during consultations [9,30], for example, by patients expressing more concerns and asking more questions [31].

Subsequently, patient participation can positively influence factors related to the quality of care, such as satisfaction with the consultation and understanding of health information provided [32]. In addition, researchers found that patient participation is related to lower anxiety [33], increased satisfaction [34-36], and improved information recall [13,37]. However, knowledge about whether and how the effects of OHIS on these outcomes are mediated by patient participation during consultation is lacking. Therefore, the aim of this study is to examine the demographic and psychosocial factors that can predict OHIS and how OHIS, in turn, influences patient outcomes via patient participation during consultations.

Predictors of OHIS

Cancer patients vary in the extent to which they seek online health information. The Comprehensive Model of Information Seeking is one of the most widely adopted models to discuss factors that could influence health information seeking [22]. In this model, demographics and psychosocial factors are seen as important determinants of how much an individual is inclined to search for health information.

Demographics

In general, studies show that demographics such as age, education level, and gender correlate with OHIS [16]. However, results are ambiguous. For example, some have shown that younger individuals seek more online health information than older individuals [16,38-40], whereas others find that older adults tend to seek more information online than their younger counterparts [41] or find no correlations with OHIS at all [42]. Frailty, or "the risk for adverse outcomes due to losses in different domains of functioning" [43], is found to be related to a decline in patients' self-management abilities, more so than chronological aging. Therefore, the level of frailty, also called biological age, might better predict a patient's ability to engage in OHIS than chronological age. In addition, several studies have shown that females seek online health information more frequently than males [16,38,40,44], whereas other studies show no associations between OHIS and gender [41,42]. With respect to education level, there is some evidence that higher educated individuals seek more online health information than lower educated individuals [44]; however, other studies show no such associations [20,42,45]. Finally, the tendency to search for health information online can also differ according to one's degree of health literacy or "the ability to perform basic reading and numerical tasks required to function in the health care environment" [46]. As described in a review study, some studies show limited evidence that people with low health literacy search less frequently for health information online, compared with people with high health literacy, whereas other studies show no differences in OHIS based on health literacy [47].

Psychosocial Factors

In addition to demographics, OHIS can also be explained by patients' psychosocial characteristics such as their degree of stress or anxiety and strategies to cope with such feelings. Higher levels of fear and anxiety in cancer patients have both been associated with the tendency to avoid cancer-related information [28,48] and with increased information needs [49]. Seeking relevant health information online might help patients to deal with the feelings of anxiety, and some patients feel relieved or comforted by the information they find online [45,50]. However, cancer patients differ in their need for cancer-related information [48], based on how they cope with a health threat. Some patients prefer only a very limited amount of information (blunting coping style), whereas others prefer as much information as possible (monitoring coping style) [51-56]. As the results are inconsistent, more research is needed, resulting in research question (RQ) 1:

- RQ 1: Are cancer patients' demographic characteristics (ie, age, gender, education level, frailty, and health literacy) and psychosocial characteristics (ie, anxiety, cancer-related

XSL•FO
RenderX

stress, and information-seeking coping style) related to OHIS?

Direct Relation of OHIS and Patient Participation

Patient Participation

OHIS may potentially better equip patients to participate in consultations with health care providers [57-59]. Actively participating in such consultations reflects patients' ability and willingness to express their needs, concerns, preferences, and expectations [32]. According to the linguistic model of patient participation in care, patients need a certain repertoire of informational resources to actively communicate during medical consultations [32]. Patients with sufficient knowledge about a topic or terminology related to the topic will discuss health issues more easily with their providers [60]. Therefore, the knowledge a patient possesses, which might be gained because of OHIS, influences a patient's ability to actively communicate and is an important factor in patient participation [29,32,61].

In addition, providing patients with an opportunity to gather information and seeking online health information can empower patients by giving them the feeling that they are better prepared for their consultations, thereby making them confident enough to actively participate during consultations [9,29]. A recent review showed that gathering online health information before a consultation resulted in patients feeling more self-assured and empowered during consultations [9].

In conclusion, seeking health information online can prepare patients for interactions with health care providers by increasing knowledge and feelings of empowerment and might, therefore, be a crucial predictor of patient participation. Therefore, we argue that more OHIS leads to greater patient participation during a consultation with a health care provider, resulting in hypothesis 1 (H1):

- H1: OHIS is positively related to cancer patients' participation during a medical consultation.

Indirect Relation of OHIS and Patient Outcomes: The Mediating Role of Patient Participation

Both OHIS and patient participation are believed to be important independent factors that influence affective and cognitive patient outcomes [6,62]. OHIS most likely influences these outcomes via patient participation because it can increase patients' illness-related knowledge and feelings of empowerment, leading to more patient participation [32]. Active patient participation can, in turn, positively affect factors that indicate quality of care [32]. Indeed, studies have found that patient participation results in less anxiety [6,33], more satisfaction [34-36], and better information recall [13,37].

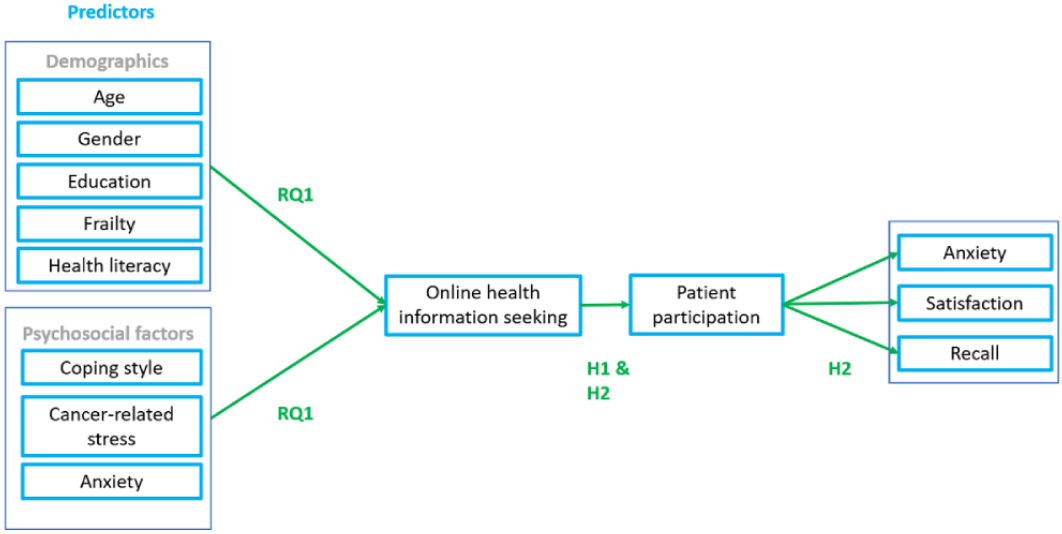
Anxiety

OHIS can positively influence emotional well-being in general, for example, by making the patient feel less stressed [5] and less anxious [10-12]. OHIS can also help patients gain knowledge about their illness [30], making them feel more empowered to discuss certain topics during consultations [9], which, in turn, can lower their stress and feelings of anxiety. If patients experience feelings of anxiety beforehand, or because of OHIS, actively participating during the consultation gives them a chance to discuss their issues with the health care provider, which might help decrease their anxiety.

On the other hand, in some cases, OHIS can increase feelings of worry and anxiety [27,63]. Patients can experience confusion because of seeking health information [27,30], which can result in feeling less comfortable to participate and act more reserved during consultations. If a patient already feels anxious because of seeking online health information and does not actively participate during consultations, the health care provider may not be able to adequately address the patient's anxiety. As a result, their anxiety may remain or increase even further. In line with this, we argue that the effect of OHIS on anxiety is mediated by patient participation during medical consultations (Figure 1), resulting in hypothesis 2a (H2a):

- H2a: Patient participation mediates the effect of OHIS on anxiety and stress after consultation.

Figure 1. Theoretical model. H1: hypothesis 1; H2: hypothesis 2; RQ1: research question 1.



XSL•FO
RenderX

Satisfaction

Generally, better informed patients are more satisfied with their health care processes [6,7,12,64]. Russ et al [8] found that patients who sought online health information were more satisfied with the information provided by the provider during a consultation when compared with patients who did not seek online health information.

A reason for this increase in satisfaction can be that seeking online health information before a consultation gives the patient a feeling of being prepared for the consultation [9]. Online information can help patients anticipate the discussion of certain topics during consultations or to consider possible treatments that will be proposed [65]. Knowing what to expect during the consultation can result in more active participation, including the expression of questions or expectations. These can subsequently be addressed by the health care provider, resulting in greater satisfaction with the consultation. In line with this, patients are more satisfied when providers are supportive of their search for online health information [66,67]. Therefore, it can be argued that OHIS leads to more satisfied patients through increased patient participation.

However, as discussed before, OHIS can also cause confusion, thereby inhibiting active patient participation. As a result, issues relevant to a patient may not be addressed, in which case the patient can feel disappointed and less satisfied with the consultation. Accordingly, research has shown that when the online findings do not match with the information discussed during consultation, for example, regarding diagnosis or treatment options, this can result in a less satisfied patient [68]. Therefore, we argue that the effect of OHIS on satisfaction with a consultation is mediated by patient participation (Figure 1), resulting in hypothesis 2b (H2b):

- H2b: Patient participation mediates the effect of OHIS on satisfaction with a consultation.

Recall

When patients engage in OHIS before a consultation and this leads to more participation during the consultation, this is likely to improve the recall of the information discussed [13,37,69-71]. One reason for the positive association between OHIS, participation, and recall is that repetition of the same information can improve information recall [72,73]. When patients search for online health information before the consultation and discuss the same information during the consultation by actively participating, this leads to a repetition in exposure to that information. In addition, exposure to a first piece of information can prime the interest for a second similar piece of information [74]. As this double exposure to the same kind of information stimulates deeper information processing, it is expected to positively influence information recall [75,76].

It can also be argued that patients who participate more actively during the consultation by asking more questions and expressing more concerns will receive more information from health care providers and are also more likely to understand the rationale and recommendations of the provider [32]. Moreover, actively participating patients are more involved and, therefore, process the information they receive during the consultations in an active manner. This active, deeper processing of information can result in better information recall [77]. Thus, we argue that the effect of OHIS on recall of the information provided during the consultation is mediated by patient participation (Figure 1), leading to hypothesis 2c (H2c):

- H2c: Patient participation mediates the effect of OHIS on information recall.

Methods

Design

A study was conducted in 6 Dutch hospitals among newly diagnosed colorectal cancer patients. All patients received the standard procedure of care provided by the hospitals without any alterations. All newly diagnosed patients who planned to undergo surgery were approached to participate in the study. Health care providers (surgeons and specialized nurses) and patients signed an informed consent form. Study participants received a consultation with a surgeon or specialized nurse in preparation for their surgery. This consultation was audio-recorded, transcribed, and content coded. Data were collected using questionnaires before and after the consultation.

This study was registered with Trialregister.nl (NTR5919) and was approved by the Review Board of the Amsterdam School of Communication Research (2017-PC-7979) and the medical ethical review boards of the hospitals that participated in the study (METC-nr: 13-061). The data collected to answer the RQs and hypotheses for this study were part of a larger investigation including multiple measurement moments.

Procedure and Participants

Participants included newly diagnosed colorectal cancer patients; those who had planned to undergo surgery, possibly in combination with other treatment and had sufficient command of the Dutch language, were able to read, and had no cognitive impairment according to their medical record (eg, dementia); and those who had provided written informed consent.

Once the consultation with the surgeon was scheduled, a specialized nurse or medical secretary asked the patients if they wanted to receive study information. Patients who agreed to being contacted about this study were approached, approximately 3 days before the consultation, by the study coordinator via phone to explain what study participation would entail. Consenting patients received additional information and the first online questionnaire at time point 1 (T1) via email. Patients were asked to complete the first questionnaire 1 day before the consultation.

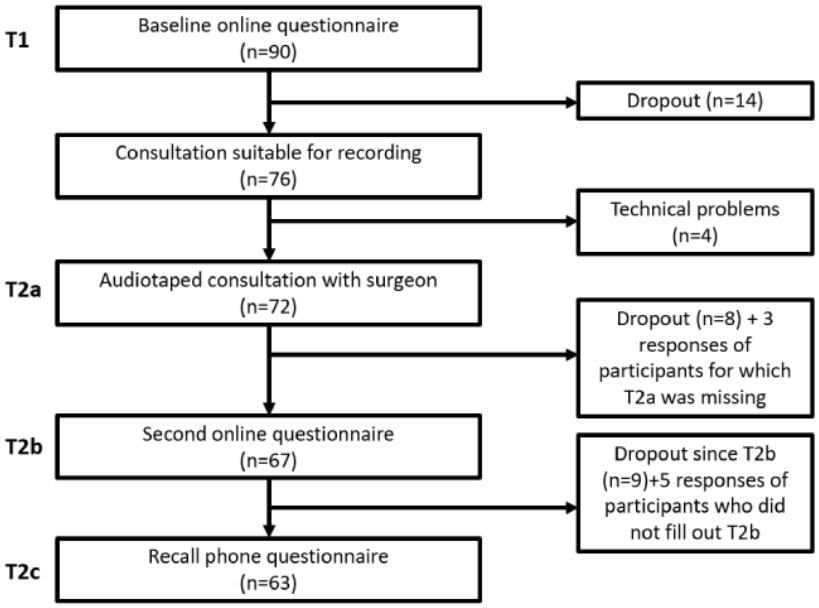
The scheduled consultation was recorded at time point 2a (T2a), and 2 days thereafter at time point 2b (T2b), the patients received the second questionnaire partly via email, including standard questions that were the same for all patients. Patients were also contacted via telephone 2 to 3 days after the consultation at time point 2c (T2c) by the research assistant or researcher to assess recall using recall questions that were tailored to the consultation.

The final sample consisted of 90 patients, as seen by 23 health care providers (surgeons and specialized nurses) in 6 Dutch

XSL•FO
RenderX

hospitals. During the study, 346 patients were reported to be suitable for participation by the specialized nurses or medical secretaries of the hospitals. A total of 285 patients were successfully approached to participate in the study. The other 61 patients either did not meet the inclusion criteria or could not be reached because of organizational or technical difficulties. Of the 285 patients who were successfully approached, 119 consented to participate in the study. As 29 of the consenting patients did not fill out the first questionnaire before the consultation, a total of 90 patients were included in the final analyses. Between the first and the following questionnaires, a number of patients dropped out, resulting in 72 consultation recordings, 67 responses on T2b, and 63 responses on T2c. More details about the dropout process are shown in Figure 2.

Figure 2. Flowchart of inclusion and dropout. T1: time point 1; T2a: time point 2a; T2b: time point 2b.



Measures T1

Demographics

Sociodemographic information was obtained in the first questionnaire with questions regarding age, gender, education level, living situation, and internet use. A total of 3 categories were formed for education level (low, middle, and high). In addition, hospital records were used to obtain medical information about diagnoses and treatments of patients.

Frailty

Frailty was measured using the Groningen Frailty Indicator [43]. This scale contains 15 items about physical functioning (mobility, multiple health problems, fatigue, and vision or hearing problems), cognitive functioning, social functioning, and psychological functioning (feelings of depression or anxiety). The total score could range from 0 to 15; however, in this study, patients scored from 0 to 11 (mean 2.80, SD 2.45), with a higher score indicating more frailty [43].

Health Literacy

A 3-item questionnaire was used to measure health literacy [46]. The items addressed one's ability to obtain and read medical information and to fill out medical forms on a 5-point scale (α=.62). The total score ranged from 1 to 5, with a higher score indicating higher health literacy (mean 4.26, SD 0.71).

Anxiety (Preconsultations)

Anxiety was measured at T1 using the short Dutch version of the State Trait Anxiety Inventory [10,78]. Patients rated the degree to which they were currently experiencing anxiety on a 4-point scale ranging from 1 (not at all) to 4 (very much). Higher scores on the scale indicate higher levels of anxiety. Patients scored on average 1.95 (SD 0.55), with scores ranging from 1 to 3.67. Cronbach alpha was good (α=.82).

Cancer-Related Stress

Cancer-related stress was measured at T1, with a subscale of the Dutch version of the Impact of Events Scale [79,80], comprising 7 items (α=.84). Participants rated the items on a 4-point Likert scale (1=*not at all*, 2=*rarely*, 3=*sometimes*, 4=*often*), with a higher score indicating higher levels of cancer-related stress. Scores ranged from 1 to 3.71, and patients scored an average of 2.03 (SD 0.70).

Coping Style

Coping style was measured using the adapted shortened version of the Threatening Medical Situation Inventory at T1 [81,82]. The scale consists of 3 items measuring monitoring intentions regarding the patients' medical situation. Items addressed intentions to (1) look for information within the threatening situation, (2) go deeply into the situation by reading about it, and (3) get information from the health care provider (α=.82). Participants responded to the statements with answer options ranging from 1 (*not at all applicable to me*) to 5 (*very much*

XSL•FO
RenderX

applicable to me) and scored an average of 3.46 (SD 1.07), with a higher score indicating higher monitoring intentions.

OHIS

On the basis of previous research [20], patients were asked to indicate on a 5-point Likert scale how often they had used the internet to seek information about their illness or treatment options before the consultation (T1). The answer options were 1 (*did not use*), 2 (*used very little*), 3 (*used sometimes*), 4 (*used regularly*), and 5 (*used often*). Patients on average scored 2.23 (SD 1.32).

Measures T2a

Patient Participation

The audiotaped consultations were transcribed and manually coded by a research assistant using 3 measures to represent patient participation. This operationalization is in line with the methods used in previous research [83-86]. First, the absolute contribution of the patient to the consultation was measured using the patient's absolute word count. Second, the relative contribution of the patient was measured by calculating the ratio of the number of words used by the patients compared with the number of words used by the health care provider. For these 2 measures, the coding process involved counting all the words used by the patient and the health care provider [83,84]. Third, the number of questions and assertions expressed by the patient during the consultation was coded using a codebook developed based on the method described by Street and Millay [32] (the complete codebook is given in Multimedia Appendix 1). A total of 10% (9/90) of the data set was double-coded by a second independent coder, resulting in acceptable intercoder reliability (κ=0.764; P<.001).

Measures T2b

Anxiety (Postconsultation)

Anxiety was measured postconsultation (T2b) in the same manner as in the preconsultation (T1). Patients on average scored 1.80 (SD 0.66). Cronbach alpha was good (α=.86).

Satisfaction With the Consultation

To measure patient satisfaction with the consultation (T2b), the 5-item *Patient Satisfaction Questionnaire* was used [87]. Items addressed the following: the extent to which the patient was satisfied in terms of needs that were met by the surgeon, if the patient felt actively involved during the consultation, the information received during the consultation, the emotional support received during the consultation, and the interaction during the consultation in general (α=.80) [84]. All the answer options ranged from 1 (*not satisfied at all*) to 5 (*completely satisfied*), and patients scored an average of 4.39 (SD 0.58).

Measures T2c

Information Recall

To measure information recall, the Netherlands Patient Information Recall Questionnaire (NPIRQ) [88] was used to compose the questions. The correct answers to the questions were (parts of) statements provided by the surgeon during the consultation. Therefore, the answers were literally derived from the transcribed consultations. Answers provided by the patients were scored as 0 (not recalled), 1 (partially recalled), and 2 (completely recalled). If the patient did not recall the information, there were 2 other answer option: "this information was not discussed" and "this information was discussed, but I can't remember the details," both resulting in a score of 0 [88].

In line with the NPIRQ guidelines, a sum score was constructed by calculating the percentage of the obtained recall score (range 6%-100%) relative to the maximum achievable score (2-26 points), with higher scores indicating better recall. Patients scored an average of 60% (SD 0.19). A total of 10% of the cases (7/63) were double-coded by 2 independent coders to check intercoder reliability (mean κ=0.71; P<.001) [89].

Statistical Analyses

The analyses are based on a 2-step process. First, the correlations between demographic and psychosocial variables and outcome variables were tested. The variables that significantly correlated with the outcome measures at a significance level of .10 were selected for follow-up analyses as control variables. Second, multivariate regression analyses were carried out to test whether demographic variables (age, gender, and education level) and psychosocial factors (frailty, coping style, stress, and anxiety before the consultation) were related to OHIS (RQ1) and if OHIS was related to patient participation (number of words used by the patient during the consultation, relative contribution a patient had in the consultation in terms of the word count ratio, and number of questions and assertions expressed; H1). For the mediation effects in H2a, H2b, and H2c, regression analyses using an SPSS macro allowing for mediation, (PROCESS model 4) [90] were conducted. In addition, to determine whether the relation between OHIS and the outcome variables differed depending on clustering within health care providers, multilevel analyses were carried out if the dependent variable correlated with health care providers [91].

Results

Sample

The age of patients included in the final analyses ranged from 39 to 88 years (mean 69.93, SD 9.93), and about two-thirds of the patients were male (59/90, 66%). Half of the patients (45/90, 50%) had a medium level of education. Patients' health literacy was relatively high (mean 4.25, SD 0.71), and they were not frail on average (mean 2.80, SD 2.45). Almost half of the patients (41/90, 46%) indicated that they did not use the internet, 12% (11/90) used the internet very little, 21% (19/90) used the internet sometimes, 16% (14/90) used the internet regularly, and 6% (5/90) used the internet often before the consultation. Nonresponse analyses revealed that participants did not differ compared with nonparticipants regarding gender ($F_{1,309}$=2.92; P=.09) but were on average significantly younger (mean 69.75, SD 9.93) than patients who did not wish to participate (mean 73.15, SD 10.30; $F_{1,297}$=7.24; P=.008). The background information of the participants is presented in Table 1.

XSL•FO
RenderX

Table 1. Sample characteristics.

Background variables[a]	Patients
Demographic information (n=90), mean (SD)	
Age (years)	69.93 (9.93)
Gender (n=90), n (%)	
Male	59 (66)
Female	31 (34)
Education level (n=88), n (%)	
Low	24 (27)
Medium	45 (51)
High	19 (22)
Health background information (n=90), mean (SD)	
Health literacy[b]	4.25 (0.71)
Frailty[c]	2.80 (2.45)
Psychosocial information (n=90), mean (SD)	
Coping style[d]	3.46 (1.07)
Online health information seeking behavior (n=90), n (%)	
Never	41 (46)
Very little	11 (12)
Sometimes	19 (21)
Regularly	14 (16)
Often	5 (6)

[a]All cells add up to 100% owing to missing data.

[b]A higher score indicates higher levels of health literacy (maximum range 1-5; reported range 1-5).

[c]A higher score indicates higher frailty (maximum range 0-15; reported range 0-11).

[d]A higher score indicates a higher information-monitoring coping style (maximum range 1-5; reported range 1-5).

Patient Participation

Recorded consultations (n=72) lasted between 4 minutes 26 seconds and 46 minutes 40 seconds, with an average duration of 20 minutes 19 seconds (SD 7.47 minutes). The number of words spoken during these consultations ranged from 488 to 6824 words (mean 2657, SD 1307.89). Patients spoke a minimum of 29 words and a maximum of 1347 words (mean 472.57, SD 295.46), whereas health care providers spoke at least 386 words and at the most 5124 words (mean 1998.83, SD 991.93). Patients scored a relative contribution to the consultation of 19.12% (472.57/2471.4) on average, ranging from 3.4% to 43.5% (SD 8.20); therefore, the ratio of health care providers ranged from 56.5% to 96.6%, with an average of 80.8% (SD 8.20).

A total of 69 patients asked at least one question, and 55 patients expressed at least one assertion. The number of questions ranged from 1 to 35 per consultation (mean 6.44, SD 6.36), and the number of assertions ranged from 1 to 10 per consultation (mean 2.30, SD 1.92). This resulted in a total number of questions and assertions ranging from 1 to 37 (mean 7.96, SD 7.03).

Predictors of OHIS

Demographics

Correlation analyses showed that age was negatively related to OHIS (r=−0.29; P=.005), suggesting that an increase in age was associated with less OHIS. Education level and OHIS were positively correlated (r=0.37; P<.001), suggesting that higher educated patients engage more in OHIS. No significant correlations were found between OHIS and gender (r=0.01; P=.91), frailty (r=−0.10; P=.35), and health literacy (r=0.15; P=.14; Table 2).

XSL•FO
RenderX

Table 2. Correlation matrix.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1. Age	—[a]																
2. Gender[b]	0.021	—															
3. Education level[c]	−0.057	0.038	—														
4. Health literacy	0.039	0.074	0.220*	—													
5. Frailty	−0.157	−0.002	−0.183	−0.295**	—												
6. Anxiety (preconsultation)	−0.286**	−0.016	−0.113	−0.041	0.461**	—											
7. Cancer-related stress	−0.294**	−0.156	−0.02	−0.045	0.203	0.554**	—										
8. Coping style	−0.205*	−0.099	0.231*	0.112	−0.115	0.013	0.198	—									
9. Health care provider	0.096	−0.172	−0.275**	−0.097	−0.162	0.119	0.109	−0.061	—								
10. consultation time	−0.043	0.079	−0.056	−0.076	−0.250*	0.088	0.06	0.127	0.509**	—							
11. Online health information seeking	−0.289**	0.012	0.369**	0.151	−0.096	0.183	0.361**	0.453**	−0.1	0.143	—						
12. Patient participation word count	−0.061	−0.229*	0.099	0.02	−0.131	0.142	0.082	0.336**	0.392**	0.525**	0.326**	—					
13. Patient participation relative contribution	−0.074	−0.103	0.086	0.077	0.111	0.039	−0.003	0.168	−0.062	−0.156	0.22	0.574**	—				
14. Patient participation questions and utterances	0.065	−0.258*	0.147	−0.076	−0.034	0.147	0.114	0.223	0.285*	0.330**	0.176	0.633**	0.295**	—			
15. Anxiety (postconsultation)	−0.067	−0.231*	−0.144	−0.124	0.435**	0.601**	0.511**	0.152	0.085	−0.031	0.238*	0.187	0.166	0.278*	—		
16. Satisfaction	0.134	−0.200	−0.174	−0.044	−0.151	−0.169	−0.121	−0.127	0.227	0.141	−0.191	0.086	−0.178	0.005	−0.360**	—	
17. Recall	−0.105	0.021	0.080	−0.126	0.073	0.150	0.161	−0.176	0.081	0.061	0.016	0.208	0.139	0.040	−0.018	0.345**	—

[a]Not applicable.

[b]Gender was dummy coded into 1=female and 2=male.

[c]Education was dummy coded into 1=low, 2=medium, and 3=high.
*P<.05, **P<.01, ***P<.001.

Psychosocial Factors

In addition, correlation analyses showed that cancer-related stress was positively correlated with OHIS (r=0.36; P<.001), implying that higher stress levels can result in more OHIS. There was a marginally significant positive correlation between anxiety before the consultation and OHIS (r=0.18; P=.08), suggesting that patients who report higher anxiety levels might engage more in OHIS. Regarding coping style, a positive correlation was found (r=0.45; P<.001), meaning patients with higher levels of monitoring coping style engaged more in OHIS (Table 2).

Regression Analyses

To test whether these variables predict OHIS, a regression analysis was conducted, including all possible predictors that significantly correlated with OHIS (age, education level, cancer-related stress, anxiety before the consultation, and coping style). The results showed that education level (B=0.54; P=.002), cancer-related stress (B=0.48; P=.02), and coping style (B=0.41;

XSL•FO
RenderX

P=.001) were positively associated with OHIS. Thus, higher educated patients, patients experiencing more cancer-related stress, and patients with higher levels of a monitoring coping style more frequently engaged in OHIS. There was no relation between age and OHIS (*B*=−0.01; *P*=.24) and between anxiety before the consultation and OHIS (*B*=0.08; *P*=.74) based on the multivariate regression. To answer RQ1, education level, cancer-related stress, and coping style are positively related to OHIS.

Relation Between OHIS and Patient Participation During the Consultation (n=71)

The correlation analyses showed that gender was significantly related to the number of words used by the patient (*r*=−0.23; *P*=.005) and the number of questions and assertions expressed by the patient (*r*=−0.26; *P*=.003), suggesting that males used fewer words and expressed fewer questions and assertions than females. Coping style was also positively related to the number of words used by the patient (*r*=0.37; *P*=.004), indicating that patients with a more monitoring coping style used more words (Table 2). There were no significant correlations between the other variables and the number of words used, the relative contribution of a patient in the consultation in terms of the word count ratio, or the number of questions and assertions expressed by the patient.

Regression analyses were carried out to test the relation between OHIS and patient participation outcomes. On the basis of the correlation analyses, gender and coping style were included as control variables for the regression analyses regarding the number of words used by the patient and gender was included as the control variable for the regression regarding the number of questions and assertions expressed. No variables were included as control variables in the regression regarding relative contribution of the patient.

Results showed OHIS was positively related to the number of words used by the patient during the consultation (*B*=50.58; *P*=.02), when controlling for gender and coping style. The relation between OHIS and the relative contribution of the patient a patient had in the consultation in terms of the word count ratio was also significant (*B*=1.99; *P*=.02). OHIS was not related to the number of questions and assertions expressed (*B*=0.74; *P*=.26), when controlling for gender. In other words, patients who engaged more in OHIS used more words during the consultation and had a larger relative contribution to the conversation but did not express more questions and assertions. Regarding H1, we can conclude that OHIS is associated with some, albeit not all, indicators of patient participation during consultations.

Relation Between OHIS and Anxiety, Satisfaction, and Recall, Mediated by Patient Participation

The correlation analyses (n=90) showed that gender (*r*=−0.23; *P*=.005), frailty (*r*=−0.44; *P*<.001), anxiety before the consultation (*r*=−0.60; *P*<.001), and cancer-related stress (*r*=−0.51; *P*<.001) were significantly related to anxiety after the consultation. Gender was also significantly related to the number of words used by the patient (*r*=−0.23; *P*=.005) and the number of questions and assertions expressed by the patient (*r*=−0.26; *P*=.003), whereas coping style was also positively related to the number of words used by the patient (*r*=0.37; *P*=.004; Table 2). These variables were included as control variables in the regression analyses regarding anxiety after the consultation. Health care provider was only significantly related to satisfaction with the information (*r*=−0.23; *P*=.005). However, multilevel analyses showed the relation between OHIS and satisfaction was not dependent on health care provider ($F_{1,4}$=−0.04; *P*=.35). There were no significant correlations between the other variables and satisfaction with the information or information recall. Therefore, no control variables were included in the regression analyses regarding satisfaction and recall.

Anxiety (n=64)

When controlling for gender, frailty, anxiety before the consultation, and cancer-related stress, OHIS was not related to anxiety after the consultation (*B*=0.07; *P*=.17). Regarding patient participation, the number of words used by the patient *B*=−0.01 *P*=.44), the relative contribution of the patient in terms of the word count ratio (*B*=0.01; *P*=.14), and the number of questions and assertions expressed by the patient (*B*=0.01; *P*=.66) were also not related to anxiety after the consultation. There was no significant mediation of OHIS on anxiety after the consultation via the number of words used by the patient, relative contribution of the patient to the consultation, or the number of questions and assertions (Table 3); thus, H2a must be rejected.

XSL•FO
RenderX

Table 3. Mediation analyses.

Relations	*B*[a]	SE	95% CI values	*t* test (*df*)	*P* value
Direct effect of OHIS[b]					
On word count[c]	68.9740	27.8861	13.1535 to 124.7945	2.4734 (5,58)	.02
On word count ratio[d]	1.9918	0.8473	0.2958 to 3.6879	2.3508 (5,58)	.02
On questions and assertions[e]	0.7349	0.6469	−0.5601 to 2.099	1.1360 (5,58)	.26
On anxiety	0.0666	0.0517	−0.0369 to 0.1701	1.2890 (8,55)	.20
On satisfaction	−0.1029	0.0560	−0.2149 to 0.0091	−0.18377 (4,59)	.07
On recall	−0.0203	0.0189	−0.0581 to 0.0175	−1.0747 (4,58)	.29
Direct effects on anxiety					
Of word count	−0.0003	0.0003	−0.0009 to 0.0004	−0.7810 (8,55)	.44
Of word count ratio	0.0141	0.0096	−0.0051 to 0.0333	1.4679 (8,55)	.15
Of questions and assertions	0.0051	0.0117	−0.0184 to 0.0286	0.4365 (8,55)	.66
Indirect effects of OHIS on anxiety					
Mediated by word count	−0.0174	0.0231	−0.0385 to 0.0549	N/A[f]	N/A
Mediated by word count ratio	0.0280	0.0219	−0.0196 to 0.0694	N/A	N/A
Mediated by questions and assertions	0.0038	0.0162	−0.0470 to 0.0225	N/A	N/A
Direct effects on satisfaction					
Of word count	0.0008	0.0004	0.0001 to 0.0015	2.2207 (4,59)	.03
Of word count ratio	−0.0223	0.0109	−0.0442 to 0.0005	−2.0487 (4,59)	.04
Of questions and assertions	−0.0087	0.0139	−0.0365 to 0.0191	−0.6246 (4,59)	.53
Indirect effects of OHIS on satisfaction					
Mediated by word count	0.0529	0.0283	0.0053 to 0.1158	N/A	N/A
Mediated by word count ratio	−0.0319	0.0254	−0.0925 to 0.0068	N/A	N/A
Mediated by questions and assertions	−0.0068	0.0162	−0.0268 to 0.0416	N/A	N/A
Direct effects on recall					
Of word count	0.0002	0.0001	0.0000 to 0.0004	1.6737 (4,58)	.10
Of word count ratio	0.0004	0.0036	−0.0068 to 0.0076	0.1033 (4,58)	.92
Of questions and assertions	−0.0025	0.0047	−0.0119 to 0.0069	−0.5359 (4,58)	.59
Indirect effects of OHIS on recall					
Mediated by word count	−0.0131	0.0091	−0.0029 to 0.0333	N/A	N/A
Mediated by word count ratio	−.0004	0.0051	−0.0092 to 0.0127	N/A	N/A
Mediated by questions and assertions	−0.0015	0.0043	−0.0084 to 0.0101	N/A	N/A

[a]*B*: Standardized β.

[b]OHIS: online health information seeking.

[c]Number of words used by the patient.

[d]Relative contribution of the patient in terms of words used by the patient compared with words used by the health care provider.

[e]Number of questions and assertions expressed by the patient.

[f]N/A: not applicable.

Satisfaction (n=64)

OHIS was marginally negatively related to satisfaction with the consultation directly (*B*=−0.10; *P*=.07), suggesting that the more a patient engaged in OHIS, the less satisfied the patient was with the consultation. The number of words used by the patient was positively related to satisfaction with the consultation (*B*=0.0008; *P*=.03), meaning the more words a patient used, the more satisfied a patient was. The relative contribution of the patient to the consultation in terms of the word count ratio was negatively related to satisfaction (*B*=−0.02; *P*=.05), suggesting that the higher the relative contribution of the patients (and

XSL•FO
RenderX

therefore automatically the lower the contribution of the health care provider), the less satisfied the patient was. There was no significant relation between the number of questions and assertions expressed by the patient and satisfaction (B=−0.01; P=.54). The indirect relation between OHIS and satisfaction, based on the number of words used by the patient, was also significant (B=0.05; 95% CI 0.0053-0.1158). This means that patients who engaged in OHIS used more words during the consultations, which, in turn, was positively related to more satisfaction with the consultation. Therefore, H2b is partly supported.

Recall

The analyses showed no significant correlation between OHIS and information recall (B=−0.02; P=.28). In addition, there was no significant relation between the number of words used by the patient (B=0.00; P=.10), the relative contribution of the patient to the consultation (B=0.01; P=.92), the number of questions and assertions expressed (r=−0.01; P=.59), and information recall. In addition, there was no significant mediation of OHIS on information recall via 1 of the patient participation measures (Table 3). This implies that H2c must be rejected.

Discussion

Review of Findings

The aim of this study is twofold. First, this study examined which demographic and psychosocial factors could predict OHIS of newly diagnosed cancer patients. Second, we investigated how OHIS subsequently relates to patient participation during consultations and how this, in turn, affects patients' anxiety, satisfaction, and information recall. Regarding demographic factors, the results showed that patients with higher levels of education were more inclined to engage in OHIS. With respect to psychosocial factors, higher levels of cancer-related stress are associated with more OHIS, and patients with a monitoring coping style also engage more in OHIS. In turn, OHIS was positively related to patient participation in terms of the number of words used by the patient during the consultation and the relative contribution of the patient in the consultation but not to the number of questions and assertions expressed.

The negative direct relation between OHIS and satisfaction shows that more OHIS leads to lower patient satisfaction. In addition, the number of words used by the patient was related to higher levels of satisfaction with the consultation, whereas the relative contribution of the patient in the consultation was related to lower levels of satisfaction. The results also showed a positive indirect relation between OHIS and satisfaction via the number of words used by the patient, meaning that patients who engaged more in OHIS used more words during the consultation, which, in turn, was positively related to satisfaction with the consultation. On the basis of these results, it can be concluded that OHIS can lead to both more and less satisfaction with the consultation, depending on the mediation of the number of words used by the patient.

Our results indicate that not all patients engage in OHIS. In particular, lower educated patients search less for health information online. This is in line with previous research in which education has been shown to positively influence OHIS [92]. Therefore, concerns raised almost 20 years ago by Lenhart et al [31,93] regarding the digital divide still appear to be valid. As our findings suggest that OHIS is related to patient participation and satisfaction with the consultation, it can be seen as problematic that a group of patients still does not engage in OHIS.

Our results show different relations between the different measures of patient participation and OHIS. First, our results seem to suggest that patients who engage in OHIS are inclined to use more words during the consultation, which, in turn, results in greater satisfaction with the consultation. This mediation may occur regardless of the reaction of the health care providers. However, satisfaction with the consultation might also be influenced by the interplay between the patient and health care provider. For example, patient participation can elicit a response in the health care provider, for example, discussing more information during consultations [94-96]. On the other hand, the health care provider may disregard the patient's input, which is more in line with studies that have shown health care providers to insufficiently meet the patient's needs [93-95]. If the relative contribution of the patient is higher, it could mean that even though the patient uses more words, the health care provider does not respond to the patient's input. This could explain why an increase in the relative contribution of the patient to the consultation is related to a decrease in satisfaction with the consultation.

Second, the undemonstrated relation between OHIS and the expression of questions and assertions contradicts previous research, suggesting that OHIS facilitates patients to express their needs and concerns [97-99]. One reason for this could be that online health content is often incorrect, incomplete, and biased [97] and is usually experienced by patients as difficult to comprehend [97-99]. If patients engage in OHIS but find information that confuses them, this might inhibit their expression of questions or assertions. In particular, if patients do not feel empowered and confident during the consultation, they might ask fewer questions and express less assertions. It might also be possible that patients did not find the right information to support them in asking questions or expressing assertions or that OHIS fulfilled patients' information needs and already answered questions patients had. This could have resulted in patients asking fewer questions during consultations. On the other hand, finding ambiguous information online could also lead to confusion resulting in patients asking more questions during the consultation. We swiftly examined the content of the transcripts to obtain a better understanding of the differences in relations between OHIS and the separate indicators of patient participation. The transcripts showed that patients who used more words but did not express more questions and assertions mostly engaged in small talk and discussed side issues unrelated to their ongoing situation. This implies that patients who are more active in OHIS are also more active during consultations in terms of using more words; however, the information they found online did not seem to empower them enough to express treatment-related questions or assertions.

XSL•FO
RenderX

We expected that OHIS would result in less anxiety after the consultation (H2a), via more patient participation, but our results did not support this. The fact that OHIS did not influence the expression of questions and assertions might explain why we also did not find an indirect relation between OHIS and anxiety via patient participation, as feelings of anxiety could not be partly dismantled by discussing them with the health care provider.

The aforementioned line of reasoning may also explain why OHIS did not lead to better information recall, indirectly via patient participation. By not expressing questions or assertions, but just talking more about other subjects, more information was added to the consultation. The amount of information this added to the consultation could have overshadowed the most important information about the diagnosis and treatment. Previous research has shown that the amount of information discussed during a consultation can negatively influence recall of the information discussed [88].

Strengths

This study is, to the best of our knowledge, the first to show a significant mediation of OHIS on satisfaction with the consultation via patient participation. Established models regarding the influence of OHIS on patient participation mainly focused on the ways in which patient participation can be increased by OHIS, for example, by increasing knowledge and feelings of empowerment [62], or how patient participation can influence patient outcomes [31,84,93]. Our findings help to connect and extend these models by linking these 2 processes together, considering both the influence of OHIS on patient participation and the relation between patient participation and patient outcomes.

A distinguishing feature of this study was the participants. Including newly diagnosed cancer patients is challenging because of the emotional burden the patients face. Therefore, another strength of this study is that we succeeded in collecting these data in a vulnerable population. The fact that this is a multicenter study, with participating patients being treated in 1 of 6 Dutch hospitals, made inclusion of the patients even harder. Although this is beneficial for the external validity of the study, differences occurred in the recruitment process between the hospitals and inclusion was more troubled in some hospitals than in others, resulting in varying inclusion rates between hospitals.

Limitations and Future Research

First, patient participation was operationalized using only quantitative measures. Therefore, we could only draw conclusions based on the quantity of patient participation and not on the quality of patient participation. Future research should also qualitatively address patient participation during consultations to gain more insight into the content of patient participation. In addition, only the utterances of the patients were analyzed. The utterances of health care providers were only included in terms of relative contribution to the consultation but not in terms of content. As it seems plausible that patients' communication is dependent on the interplay between the partakers in that consultation [31,84,92], it is advisable to analyze the behavior of all parties taking part in the consultation in future research. In addition, only behavioral measures were used in this study to measure patient participation. Adding measures of perceived participation would be a valuable addition and is, therefore, recommended for future research.

A limitation that could have influenced the relations with information recall is that in this study, the number of recall questions was based on the amount of information the patient received from the health care provider during the consultation. This means that the more information was provided, the more recall questions the patient had to answer. The amount of information is known to be negatively related to the ability to correctly recall this information [100,101], and a higher number of questions can mean a higher chance of making mistakes. The researchers of this study deliberately chose to tailor the recall questions to the consultations of each separate patient because asking a fixed set of recall questions meant asking questions about topics that were not discussed with the patient, which was seen as unethical. Researchers can decide on asking a maximum number of questions per topic in the case of long consultations.

Finally, as our results show that OHIS does not lead to expressing questions or utterances, we encourage researchers to further investigate the effects of other types of online health information, such as online tools specifically developed and offered to patients. Previous research has shown that online health information developed and offered to a specific patient population, including preparatory tools such as question prompt lists or information tailored to a patient's situation, can be effective in increasing patient participation [99,100].

Practically, as we see a relation between some measures of patient participation and satisfaction, but not all, this study shows the importance of providing patients with the right tools to search for online health information that stimulates participation by means of expressing questions and utterances during consultations. In particular, because OHIS can also increase worry and confusion [27,30,63], health care providers are advised to guide patients with clear instructions on how to search for information online. For example, hospitals could provide patients with flyers, including information about which websites are reliable and which websites are not.

Conclusions

This study showed that younger patients, higher educated patients, patients who experience more cancer-related stress, and patients with a monitoring coping style are more likely to engage in OHIS. OHIS is positively related to the patient's absolute contribution during a consultation, which, in turn, results in the patient being more satisfied with the consultation. The results are an important addition to established models regarding the influence of OHIS.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all health care providers who participated in this study: Dr Bosker, Dr Liem, Dr Talsma, and Dr Torrenga from Deventer Hospital; Dr Steup, Dr Brouwers, Dr Boutkan, Dr Bastiaanse, and Dr Gooike from Haga Hospital; Dr Welten, Dr Hulsewe, Dr van Bastelaar, Dr Stoot, Dr Jansen, Dr Teuwen, Dr Sosef, and Dr Van Esser from Zuyderland Medical Center; Ms Koet and Ms Hoogendoorn from Albert Schweitzer Hospital; and Dr Fransen from Laurentius Hospital. In addition, the authors would like to thank all research assistants who helped with the data collection: Romy Koopsen, Chrissy Ooteman, Sanne Tamboer, Lisa van der Burg, Monique Alblas, Esther van der Woude, Manon Eggink, Laura Welbergen, Rodienne Kramers, Hasina Nezami, Jeresa Willems, Lisa Numeijer, Nikki Sunnen, Amber Sibbing, Max Tanaka, Nina Mollee, Hang Nguyen, Yayla Ilksoy, Maartje van Will, Nena van Ruiten, Sabine Bijvoet, Job Schipper, and Larissa de Looper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook of patient participation.
[PDF File (Adobe PDF File), 159 KB - jmir_v23i7e23670_app1.pdf]

References

1. McHugh SM, Corrigan M, Morney N, Sheikh A, Lehane E, Hill ADK. A quantitative assessment of changing trends in internet usage for cancer information. World J Surg 2011 Feb 23;35(2):253-257. [doi: 10.1007/s00268-010-0830-8] [Medline: 20972679]
2. Basch EM, Thaler HT, Shi W, Yakren S, Schrag D. Use of information resources by patients with cancer and their companions. Cancer 2004 Jun 01;100(11):2476-2483 [FREE Full text] [doi: 10.1002/cncr.20261] [Medline: 15160355]
3. Castleton K, Fong T, Wang-Gillam A, Waqar MA, Jeffe DB, Kehlenbrink L, et al. A survey of internet utilization among patients with cancer. Support Care Cancer 2011 Aug 18;19(8):1183-1190. [doi: 10.1007/s00520-010-0935-5] [Medline: 20556435]
4. Barnes LLB, Khojasteh JJ, Wheeler D. Cancer information seeking and scanning: sources and patterns. Health Educ J 2017 Oct 25;76(7):853-868. [doi: 10.1177/0017896917717542]
5. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation study. Health Commun 2017 Aug;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
6. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
7. Weiss SM, Wengert PA, Martinez EM, Sewall W, Kopp ERN. Patient satisfaction with decision-making for breast cancer therapy. Ann Surg Oncol 1996 May;3(3):285-289 [FREE Full text] [doi: 10.1007/bf02306284]
8. Russ H, Giveon SM, Catarivas MG, Yaphe J. The effect of the Internet on the patient-doctor relationship from the patient's perspective: a survey from primary care. Isr Med Assoc J 2011 Apr;13(4):220-224 [FREE Full text] [Medline: 21598810]
9. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
10. Marteau TM, Kidd J, Cuddeford L, Walker P. Reducing anxiety in women referred for colposcopy using an information booklet. Br J Health Psychol 1996;1(12):181-189. [doi: 10.1111/j.2044-8287.1996.tb00501.x]
11. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
12. Nguyen MH, Smets EMA, Bol N, Bronner MB, Tytgat KMAJ, Loos EF, et al. Fear and forget: how anxiety impacts information recall in newly diagnosed cancer patients visiting a fast-track clinic. Acta Oncologica 2018 Sep 28;58(2):182-188. [doi: 10.1080/0284186x.2018.1512156]
13. Richard C, Glaser E, Lussier MT. Communication and patient participation influencing patient recall of treatment discussions. Health Expect 2017 Aug 21;20(4):760-770 [FREE Full text] [doi: 10.1111/hex.12515] [Medline: 27868327]
14. Feathers A, Yen T, Yun L, Strizich G, Swaminath A. Internet searches about therapies do not impact willingness to accept prescribed therapy in inflammatory bowel disease patients. Dig Dis Sci 2016 Apr 10;61(4):1013-1020. [doi: 10.1007/s10620-015-3981-5] [Medline: 26660681]
15. Abdul-Muhsin H, Tyson M, Raghu S, Humphreys M. The informed patient: an analysis of information seeking behavior and surgical outcomes among men with benign prostatic hyperplasia. Am J Mens Health 2017 Jan 08;11(1):147-153 [FREE Full text] [doi: 10.1177/1557988315617527] [Medline: 26614445]
16. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2018-12-23] [WebCite Cache ID 74sbsnhYG]

XSL•FO
RenderX

17. Johnson JD, Donohue WA, Atkin CK, Johnson S. A comprehensive model of information seeking. Sci Commun 2016 Aug 18;16(3):274-303. [doi: 10.1177/1075547095016003003]
18. Johnson JD. Factors Distinguishing Regular Readers of Breast Cancer Information in Magazines. Women & Health 1997 Sep;26(1):7-27. [doi: 10.1300/j013v26n01_02]
19. Posma ER, van Weert JCM, Jansen J, Bensing JM. Older cancer patients' information and support needs surrounding treatment: an evaluation through the eyes of patients, relatives and professionals. BMC Nurs 2009 Jan 19;8:1 [FREE Full text] [doi: 10.1186/1472-6955-8-1] [Medline: 19152675]
20. Muusses LD, van Weert JCM, van Dulmen S, Jansen J. Chemotherapy and information-seeking behaviour: characteristics of patients using mass-media information sources. Psychooncology 2012 Sep 13;21(9):993-1002. [doi: 10.1002/pon.1997] [Medline: 21671305]
21. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
22. Johnson JD, Meischke H. A comprehensive model of cancer-related information seeking applied to magazines. Human Comm Res 1993 Mar;19(3):343-367. [doi: 10.1111/j.1468-2958.1993.tb00305.x]
23. Hartoonian N, Ormseth SR, Hanson ER, Bantum EO, Owen JE. Information-seeking in cancer survivors: application of the comprehensive model of information seeking to HINTS 2007 data. J Health Commun 2014 Apr 17;19(11):1308-1325. [doi: 10.1080/10810730.2013.872730] [Medline: 24742287]
24. Van Stee SK, Yang Q. Online Cancer Information Seeking: Applying and Extending the Comprehensive Model of Information Seeking. Health Commun 2018 Dec 30;33(12):1583-1592. [doi: 10.1080/10410236.2017.1384350] [Medline: 29083231]
25. Lee CJ, Zhao X, Pena-y-Lillo M. Theorizing the pathways from seeking and scanning to mammography screening. Health Commun 2016 Jun 18;31(1):117-128. [doi: 10.1080/10410236.2014.942769] [Medline: 26086195]
26. Sanders R, Linn AJ. A mixed method study investigating the impact of talking about patients' internet use on patient-reported outcomes. J Health Commun 2018;23(9):815-823. [doi: 10.1080/10810730.2018.1514443] [Medline: 30351205]
27. Linn AJ, van Weert JCM, Gebeyehu BG, Sanders R, Diviani N, Smit EG, et al. Patients' online information-seeking behavior throughout treatment: the impact on medication beliefs and medication adherence. Health Commun 2019 Nov 27;34(12):1461-1468. [doi: 10.1080/10410236.2018.1500430] [Medline: 30052088]
28. Bol N, Linn AJ, Smets EMA, Verdam MGE, van Weert JCM. Tailored communication for older patients with cancer: Using cluster analysis to identify patient profiles based on information needs. J Geriatr Oncol 2020 Jul;11(6):944-950 [FREE Full text] [doi: 10.1016/j.jgo.2020.01.004] [Medline: 31974067]
29. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
30. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. J Med Syst 2003 Feb;27(1):67-84. [doi: 10.1023/a:1021061229743] [Medline: 12617199]
31. Street RL, Gordon HS, Ward MM, Krupat E, Kravitz RL. Patient participation in medical consultations: why some patients are more involved than others. Med Care 2005 Oct;43(10):960-969. [doi: 10.1097/01.mlr.0000178172.40344.70] [Medline: 16166865]
32. Street RLJ, Millay B. Analyzing patient participation in medical encounters. Health Commun 2001;13(1):61-73. [doi: 10.1207/S15327027HC1301_06] [Medline: 11370924]
33. Clayman ML, Bylund CL, Chewning B, Makoul G. The impact of patient participation in health decisions within medical encounters. Med Decis Making 2015 Nov 19;36(4):427-452. [doi: 10.1177/0272989x15613530]
34. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. Soc Sci Med 2003 Sep;57(5):791-806. [doi: 10.1016/s0277-9536(02)00449-5] [Medline: 12850107]
35. Epstein RM, Street RL. Patient-centered communication in cancer care: promoting healing and reducing suffering. National Cancer Institute. 2007. URL: https://cancercontrol.cancer.gov/sites/default/files/2020-06/pcc_monograph.pdf [accessed 2021-03-01]
36. Jackson JL. Communication about symptoms in primary care: impact on patient outcomes. J Altern Complement Med 2005 Dec;11 Suppl 1(supplement 1):S51-S56. [doi: 10.1089/acm.2005.11.s-51] [Medline: 16332187]
37. Dillon PJ. Assessing the influence of patient participation in primary care medical interviews on recall of treatment recommendations. Health Commun 2012 Jan;27(1):58-65. [doi: 10.1080/10410236.2011.569000] [Medline: 21707391]
38. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128:151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
39. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
40. van Weert JCM, Bolle S, Muusses LD. Age and age-related differences in internet usage of cancer patients. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LNCS(PART 3). Switzerland: Springer; 2014:403-414.

XSL•FO
RenderX

41. Rice RE. Influences, usage, and outcomes of internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
42. Renahy E, Parizot I, Chauvin P. Determinants of the frequency of online health information seeking: results of a web-based survey conducted in France in 2007. Inform Health Soc Care 2010 Jan 19;35(1):25-39 [FREE Full text] [doi: 10.3109/17538150903358784] [Medline: 20302437]
43. Schuurmans H, Steverink N, Lindenberg S, Frieswijk N, Slaets JPJ. Old or frail: what tells us more? J Gerontol A Biol Sci Med Sci 2004 Sep 01;59(9):962-965. [doi: 10.1093/gerona/59.9.m962] [Medline: 15472162]
44. Myrick JG, Willoughby JF. Educated but anxious: how emotional states and education levels combine to influence online health information seeking. Health Informatics J 2019 Sep 21;25(3):649-660 [FREE Full text] [doi: 10.1177/1460458217719561] [Medline: 28728457]
45. Lorence DP, Park H, Fox S. Assessing health consumerism on the web: a demographic profile of information-seeking behaviors. J Med Syst 2006 Aug 29;30(4):251-258. [doi: 10.1007/s10916-005-9004-x] [Medline: 16978004]
46. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
47. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. Patient Educ Couns 2011 Mar;82(3):346-354. [doi: 10.1016/j.pec.2010.12.008] [Medline: 21239134]
48. Leydon GM, Boulton M, Moynihan C, Jones A, Mossman J, Boudioni M, et al. Cancer patients' information needs and information seeking behaviour: in depth interview study. Br Med J 2000 Apr 01;320(7239):909-913 [FREE Full text] [doi: 10.1136/bmj.320.7239.909] [Medline: 10742000]
49. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
50. Fox S. Online health search 2006. Pew Research Center. 2006. URL: https://www.pewresearch.org/internet/2006/10/29/online-health-search-2006/ [accessed 2021-03-01]
51. Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. J Pers Soc Psychol 1987;52(2):345-353. [doi: 10.1037/0022-3514.52.2.345]
52. Rood JAJ, Van Zuuren FJ, Stam F, van der Ploeg T, Huijgens PC, Verdonck-de Leeuw IM. Cognitive coping style (monitoring and blunting) and the need for information, information satisfaction and shared decision making among patients with haematological malignancies. Psychooncology 2015 May 21;24(5):564-571. [doi: 10.1002/pon.3699] [Medline: 25333875]
53. Timmermans LM, van der Maazen RWM, Leer JWH, Kraaimaat FW. Palliative or curative treatment intent affects communication in radiation therapy consultations. Psychooncology 2006 Aug;15(8):713-725. [doi: 10.1002/pon.1008] [Medline: 16308906]
54. Ong LML, Visser MR, Van Zuuren FJ, Rietbroek RC, Lammes FB, De Haes JC. Cancer patients' coping styles and doctor-patient communication. Psycho-Oncology 2001 Mar;8(2):155-166. [doi: 10.1002/(sici)1099-1611(199903/04)8:2<155::aid-pon350>3.0.co;2-a]
55. Muris P, van Zuuren FJ, de Vries S. Monitoring, blunting and situational anxiety: a laboratory study on coping with a quasi-medical stressor. Pers Individ Differ 1994 Mar;16(3):365-372. [doi: 10.1016/0191-8869(94)90061-2]
56. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Libr Assoc 2005 Jul;93(3):353-362 [FREE Full text] [Medline: 16059425]
57. Dedding C, van Doorn R, Winkler L, Reis R. How will e-health affect patient participation in the clinic? A review of e-health studies and the current evidence for changes in the relationship between medical professionals and patients. Soc Sci Med 2011 Jan;72(1):49-53. [doi: 10.1016/j.socscimed.2010.10.017] [Medline: 21129832]
58. Hashimoto H, Fukuhara S. The influence of locus of control on preferences for information and decision making. Patient Educ Couns 2004 Nov;55(2):236-240. [doi: 10.1016/j.pec.2003.09.010] [Medline: 15530760]
59. Shuyler KS, Knight KM. What are patients seeking when they turn to the internet? Qualitative content analysis of questions asked by visitors to an orthopaedics Web site. J Med Internet Res 2003 Oct 10;5(4):e24 [FREE Full text] [doi: 10.2196/jmir.5.4.e24] [Medline: 14713652]
60. Street RL, Voigt B, Geyer C, Manning T, Swanson GP. Increasing patient involvement in choosing treatment for early breast cancer. Cancer 1995 Dec 01;76(11):2275-2285. [doi: 10.1002/1097-0142(19951201)76:11<2275::aid-cncr2820761115>3.0.co;2-s]
61. Sahlsten MJM, Larsson IE, Sjöström B, Plos KAE. An analysis of the concept of patient participation. Nurs Forum 2008 Jan;43(1):2-11. [doi: 10.1111/j.1744-6198.2008.00090.x] [Medline: 18269439]
62. Street RLJ. Communication in medical encounters: an ecological perspective. In: The Routledge Handbook of Health Communication. Oxfordshire, United Kingdom: Routledge; 2003:A.
63. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
64. Fogel J, Albert SM, Schnabel F, Ditkoff BA, Neugut AI. Internet use and social support in women with breast cancer. Health Psychol 2002;21(4):398-404. [doi: 10.1037/0278-6133.21.4.398]

XSL•FO
RenderX

65. Czaja R, Manfredi C, Price J. The determinants and consequences of information seeking among cancer patients. J Health Commun 2003 Nov;8(6):529-562. [doi: 10.1080/716100418]
66. Bylund CL, Gueguen JA, D'Agostino TA, Li Y, Sonet E. Doctor-patient communication about cancer-related internet information. J Psychosoc Oncol 2010 Feb 22;28(2):127-142 [FREE Full text] [doi: 10.1080/07347330903570495] [Medline: 20391071]
67. Hay MC, Cadigan RJ, Khanna D, Strathmann C, Lieber E, Altman R, et al. Prepared patients: internet information seeking by new rheumatology patients. Arthritis Rheum 2008 Apr 15;59(4):575-582 [FREE Full text] [doi: 10.1002/art.23533] [Medline: 18383399]
68. Ahluwalia S, Murray E, Stevenson F, Kerr C, Burns J. 'A heartbeat moment': qualitative study of GP views of patients bringing health information from the internet to a consultation. Br J Gen Pract 2010 Feb 01;60(571):88-94. [doi: 10.3399/bjgp10x483120]
69. Harrington J. Improving patients' communication with doctors: a systematic review of intervention studies. Patient Educ Couns 2004 Jan;52(1):7-16. [doi: 10.1016/s0738-3991(03)00017-x]
70. van Weert JCM, Jansen J, Spreeuwenberg PM, van Dulmen S, Bensing JM. Effects of communication skills training and a Question Prompt Sheet to improve communication with older cancer patients: a randomized controlled trial. Crit Rev Oncol Hematol 2011 Oct;80(1):145-159. [doi: 10.1016/j.critrevonc.2010.10.010] [Medline: 21075644]
71. Sansoni JE, Grootemaat P, Duncan C. Question Prompt Lists in health consultations: a review. Patient Educ Couns 2015 Jun 03;98(12):1454-1464. [doi: 10.1016/j.pec.2015.05.015] [Medline: 26104993]
72. Kreps GL. Online information and communication systems to enhance health outcomes through communication convergence. Hum Commun Res 2017 Jun 30;43(4):518-530. [doi: 10.1111/hcre.12117]
73. Watson PWB, McKinstry B. A systematic review of interventions to improve recall of medical advice in healthcare consultations. J R Soc Med 2009 Jun;102(6):235-243 [FREE Full text] [doi: 10.1258/jrsm.2009.090013] [Medline: 19531618]
74. Voorveld HAM, Neijens PC, Smit EG. Opening the black box: Understanding cross-media effects. J Market Comm 2011 Apr;17(2):69-85. [doi: 10.1080/13527260903160460]
75. Dijkstra M. An experimental investigation of synergy effects in multiple-media advertising campaigns. 2002. URL: https://research.tilburguniversity.edu/en/publications/an-experimental-investigation-of-synergy-effects-in-multiple-medi [accessed 2021-03-01]
76. Epstein S. Integration of the cognitive and the psychodynamic unconscious. Am Psychol 1994;49(8):709-724. [doi: 10.1037/0003-066x.49.8.709]
77. Petty RE, Cacioppo JT. The elaboration likelihood model. In: Advances in Experimental Social Psychology. Amsterdam: Elsevier; 1986:123-205.
78. van der Bij AK, de Weerd S, Cikot RJ, Steegers EA, Braspenning JC. Validation of the dutch short form of the state scale of the Spielberger State-Trait Anxiety Inventory: considerations for usage in screening outcomes. Community Genet 2003 Oct 17;6(2):84-87. [doi: 10.1159/000073003] [Medline: 14560068]
79. Pieterse A, van Dulmen S, Ausems M, Schoemaker A, Beemer F, Bensing J. QUOTE-gene(ca): development of a counselee-centered instrument to measure needs and preferences in genetic counseling for hereditary cancer. Psychooncology 2005 May;14(5):361-375. [doi: 10.1002/pon.853] [Medline: 15386761]
80. van der Ploeg E, Mooren TTM, Kleber RJ, van der Velden PG, Brom D. Construct validation of the Dutch version of the impact of event scale. Psychol Assess 2004 Mar;16(1):16-26. [doi: 10.1037/1040-3590.16.1.16] [Medline: 15023089]
81. Bronner MB, Nguyen MH, Smets EM, van de Ven AWH, van Weert JCM. Anxiety during cancer diagnosis: Examining the influence of monitoring coping style and treatment plan. Psychooncology 2018 Feb 26;27(2):661-667. [doi: 10.1002/pon.4560] [Medline: 28976610]
82. van Zuuren FJ, de Groot KI, Mulder NL, Peter M. Coping with medical threat: an evaluation of the Threatening Medical Situations Inventory (TMSI). Pers Individ Differ 1996 Jul;21(1):21-31. [doi: 10.1016/0191-8869(96)00029-3]
83. Schinkel S, Van Weert JCM, Kester JAM, Smit EG, Schouten BC. Does Media Use Result in More Active Communicators? Differences Between Native Dutch and Turkish-Dutch Patients in Information-Seeking Behavior and Participation During Consultations With General Practitioners. J Health Commun 2015 Aug 14;20(8):910-919. [doi: 10.1080/10810730.2015.1018600] [Medline: 26073918]
84. Zandbelt LC, Smets EM, Oort FJ, Godfried MH, de Haes HC. Patient participation in the medical specialist encounter: does physicians' patient-centred communication matter? Patient Educ Couns 2007 Mar;65(3):396-406. [doi: 10.1016/j.pec.2006.09.011] [Medline: 17085006]
85. Ford S, Hall A, Ratcliffe D, Fallowfield L. The Medical Interaction Process System (MIPS): an instrument for analysing interviews of oncologists and patients with cancer. Soc Sci Med 2000 Feb;50(4):553-566. [doi: 10.1016/s0277-9536(99)00308-1]
86. Russell RL, Stiles WB. Categories for classifying language in psychotherapy. Psychol Bull 1979;86(2):404-419. [doi: 10.1037/0033-2909.86.2.404]
87. Blanchard CG, Ruckdeschel JC, Fletcher BA, Blanchard EB. The impact of oncologists' behaviors on patient satisfaction with morning rounds. Cancer 1986 Jul 15;58(2):387-393. [doi: 10.1002/1097-0142(19860715)58:2<387::aid-cncr2820580233>3.0.co;2-3]

XSL•FO
RenderX

88. Jansen J, Butow PN, van Weert JCM, van Dulmen S, Devine RJ, Heeren TJ, et al. Does age really matter? Recall of information presented to newly referred patients with cancer. J Clin Oncol 2008 Nov 20;26(33):5450-5457. [doi: 10.1200/JCO.2007.15.2322] [Medline: 18936478]
89. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007 Apr;1(1):77-89. [doi: 10.1080/19312450709336664]
90. Hayes A. Integrating mediation and moderation analysis: fundamentals using PROCESS. URL: https://statisticalhorizons.com/seminars/public-seminars/medmodlecture [accessed 2021-03-01]
91. Hox JJ, Moerbeek M, van de Schoot R. Multilevel analysis: techniques and applications. New York: Routledge; 2017.
92. Lenhart A, Horrigan JB, Horrigan L. Re-visualizing the digital divide as a digital spectrum (2003). IT & Society. 2003. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.177.1405 [accessed 2021-03-01]
93. Cegala DJ, Street RL, Clinch CR. The impact of patient participation on physicians' information provision during a primary care medical interview. Health Commun 2007 May 22;21(2):177-185. [doi: 10.1080/10410230701307824] [Medline: 17523863]
94. Fallowfield L, Jenkins V. Current concepts of communication skills training in oncology. Recent Results Cancer Res 2006;168:105-112. [doi: 10.1007/3-540-30758-3_10] [Medline: 17073196]
95. Hack TF, Degner LF, Parker PA, SCRN Communication Team. The communication goals and needs of cancer patients: a review. Psychooncology 2005 Oct;14(10):831-845. [doi: 10.1002/pon.949] [Medline: 16200519]
96. Puts MTE, Papoutsis A, Springall E, Tourangeau AE. A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. Support Care Cancer 2012 Jul;20(7):1377-1394. [doi: 10.1007/s00520-012-1450-7] [Medline: 22476399]
97. Diviani N, van den Putte B, Giani S, van Weert JCM. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
98. Flanagin AJ, Metzger MJ. Perceptions of internet information credibility. Journal Mass Commun Q 2016 Jun 25;77(3):515-540. [doi: 10.1177/107769900007700304]
99. Langille M, Bernard A, Rodgers C, Hughes S, Leddin D, van Zanten SV. Systematic review of the quality of patient information on the internet regarding inflammatory bowel disease treatments. Clin Gastroenterol Hepatol 2010 Apr;8(4):322-328. [doi: 10.1016/j.cgh.2009.12.024] [Medline: 20060070]
100. Brandes K, Linn AJ, Butow PN, van Weert JCM. The characteristics and effectiveness of Question Prompt List interventions in oncology: a systematic review of the literature. Psychooncology 2015 Mar 31;24(3):245-252. [doi: 10.1002/pon.3637] [Medline: 25082386]
101. Epstein RM, Duberstein PR, Fenton JJ, Fiscella K, Hoerger M, Tancredi DJ, et al. Effect of a patient-centered communication intervention on oncologist-patient communication, quality of life, and health care utilization in advanced cancer: the VOICE randomized clinical trial. JAMA Oncol 2017 Jan 01;3(1):92-100. [doi: 10.1001/jamaoncol.2016.4373] [Medline: 27612178]

Abbreviations

H1: Hypothesis 1
H2a: Hypothesis 2a
H2b: Hypothesis 2b
H2c: Hypothesis 2c
NPIRQ: Netherlands Patient Information Recall Questionnaire
OHIS: online health information seeking
RQ: research question
T1: Time point 1
T2a: Time point 2a
T2b: Time point 2b
T2c: Time point 2c

XSL•FO
RenderX

Edited by R Kukafka; submitted 19.08.20; peer-reviewed by R Street, S LaValley; comments to author 10.10.20; revised version received 04.12.20; accepted 31.01.21; published 05.07.21.

Please cite as:
de Looper M, van Weert JCM, Schouten BC, Bolle S, Belgers EHJ, Eddes EH, Smets EMA
The Influence of Online Health Information Seeking Before a Consultation on Anxiety, Satisfaction, and Information Recall, Mediated by Patient Participation: Field Study
J Med Internet Res 2021;23(7):e23670
URL: https://www.jmir.org/2021/7/e23670
doi:10.2196/23670
PMID:34255657

©Melanie de Looper, Julia C M van Weert, Barbara C Schouten, Sifra Bolle, Eric H J Belgers, Eric H Eddes, Ellen M A Smets. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

How Patient Work Changes Over Time for People With Multimorbid Type 2 Diabetes: Qualitative Study

Kathleen Yin[1], BPharm, PhD; Joshua Jung[1], BClinSC; Enrico Coiera[1], MBBS, PhD; Kenneth W K Ho[2], MBBS, PhD; Sanjyot Vagholkar[2], MBBS, MPH, PhD; Ann Blandford[3], MA, PhD; Frances Rapport[4], BA, MPhil, PhD; Annie Y S Lau[1], BE, PhD

[1]Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, North Ryde, Australia

[2]Faculty of Medicine, Health, and Human Sciences, Macquarie University, Sydney, Australia

[3]UCL Interaction Centre, University College London, London, United Kingdom

[4]Centre for Healthcare Resilience and Implementation Science, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Corresponding Author:
Kathleen Yin, BPharm, PhD
Centre for Health Informatics
Australian Institute of Health Innovation
Macquarie University
Level 6
75 Talavera Road
North Ryde, 2113
Australia
Phone: 61 2 9850 2477
Email: kathleen.yin@mq.edu.au

Abstract

Background: The experiences of patients change throughout their illness trajectory and differ according to their medical history, but digital support tools are often designed for one specific moment in time and do not change with the patient as their health state changes. This presents a fragmented support pattern where patients have to move from one app to another as they move between health states, and some subpopulations of patients do not have their needs addressed at all.

Objective: This study aims to investigate how patient work evolves over time for those living with type 2 diabetes mellitus and chronic multimorbidity, and explore the implications for digital support system design.

Methods: In total, 26 patients with type 2 diabetes mellitus and chronic multimorbidity were recruited. Each interview was conducted twice, and interviews were transcribed and analyzed according to the Chronic Illness Trajectory Model.

Results: Four unique illness trajectories were identified with different patient work goals and needs: living with stable chronic conditions involves patients seeking to make patient work as routinized and invisible as possible; dealing with cycles of acute or crisis episodes included heavily multimorbid patients who sought support with therapy adherence; responding to unstable changes described patients currently experiencing rapid health changes and increasing patient work intensity; and coming back from crisis focused on patients coping with a loss of normalcy.

Conclusions: Patient work changes over time based on the experiences of the individual, and its timing and trajectory need to be considered when designing digital support interventions.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-022163

(J Med Internet Res 2021;23(7):e25992) doi:10.2196/25992

KEYWORDS

patient work; self-management; diabetes; chronic conditions; chronic illness trajectory; consumer informatics

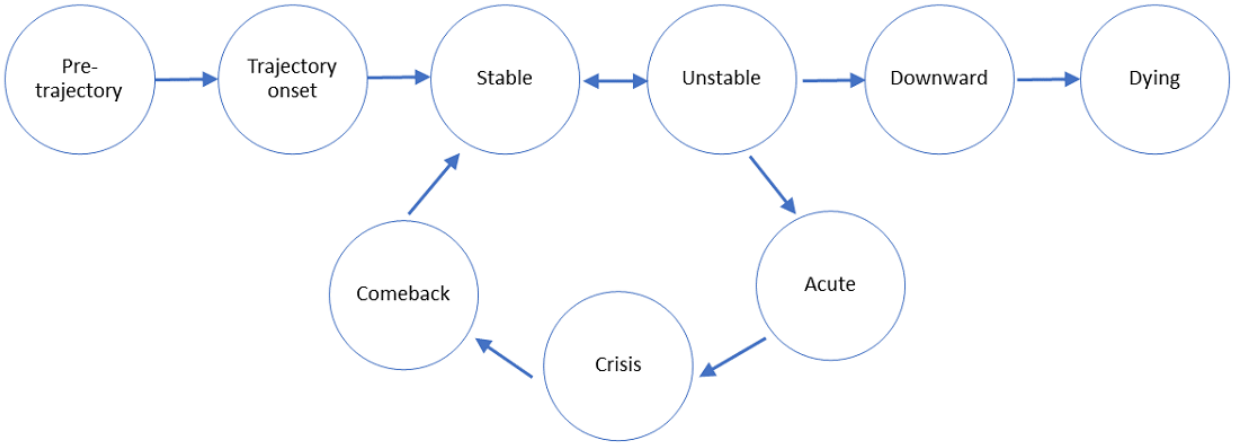
XSL•FO
RenderX

Introduction

Background and Significance

Patient work is defined as health-related tasks and actions a patient undertakes in their self-management of health conditions [1,2]. Initially proposed by Corbin and Strauss [3], the concept recognizes that patients conduct a variety of such actions, with tasks ranging from physical to cognitive, ranging from visible to invisible, conducted alone, or requiring assistance from others [4]. The Chronic Illness Trajectory Model [5], also created by Corbin and Strauss, describes how the course of illness changes over time. Patients may shift between different illness phases repeatedly over their lifetime, as their conditions fluctuate (Figure 1 [5]). This model has been used to describe patients with conditions such as injury rehabilitation [6], metastatic cancer [7], poststroke recovery [8], and multiple sclerosis [9], with evidence suggesting that patient work needs and goals change as participants move between phases [6-8].

Figure 1. The Chronic Illness Trajectory Model.



However, current digital technologies for patients with chronic disease struggle to incorporate changing health needs and goals. For example, newly diagnosed patients with type 2 diabetes prefer information on lifestyle alteration and available treatment [10,11], whereas patients with long-term complications, such as established diabetic retinopathy, require strict symptom monitoring [12,13]. Specific subpopulations have significant differences in health care preferences and goals, requiring varied information and recommendations. Studies that assessed targeted subpopulations have revealed that patients at different stages of their health condition have different information preferences [11] and that the use of apps is heavily influenced by contextual factors [14]. As such, patients experience *growing out* of the apps they took up early in their diagnosis, taking up new apps as they find their health care goals changing over the course of the illness.

Digital systems that are sensitive to the changing needs of their users would be able to help people keep using the same app over time, reducing the need to seek out more appropriate apps and ensuring continued tracking of health care data over time, especially for those who have comorbidities affecting multiple aspects of their health.

Therefore, gaps exist in our understanding of how patient work tasks and patient needs change over the phases of the chronic illness trajectory and how digital health apps can be improved by being designed with such changes in mind. We contribute to solving this issue by identifying the types of tasks patients engage in at different phases of their type 2 diabetes and the types of trajectories patients may experience over time and providing suggestions on how digital interventions could be designed to detect or anticipate changes in illness phases to provide maximum support.

Objective

In this paper, we examine how patient work tasks and goals change over the chronic illness trajectory, focusing on multimorbid patients with type 2 diabetes mellitus self-managing in the community.

Methods

Overview

We undertook interviews of multimorbid community-dwelling patients with type 2 diabetes mellitus, as described in detail in the complete study protocol [15]. Ethics approval was obtained from the Macquarie University Human Research Ethics Committee for Medical Sciences (reference number 5201700718).

Recruitment

Participants were recruited purposively from endocrinology clinics across metropolitan Sydney. Inclusion criteria were fluency in English, diagnosis of type 2 diabetes mellitus with at least one chronic comorbidity, and ability to legally provide consent.

Potential participants were suggested by clinicians, followed by researchers approaching these patients with a telephone call, explaining the purpose of the study and the processes involved. The researchers then sent a study information pamphlet and consent form via email or mail. If the person agreed to participate, the researcher arranged a time and location for the

XSL•FO
RenderX

interviews. The researcher did not disclose their professional backgrounds.

In total, 26 participants were interviewed twice. A total of 52 individuals were approached during the recruitment. From the pool of 52 people, 5 (10%) did not meet the selection criteria and were excluded (2/52, 4% did not have comorbidities and 3/52, 6% did not speak English to the required standard), 6 (12%) agreed to participate but later withdrew, and 15 (29%) declined the invitation or were unable to participate.

Data Collection

Each participant was interviewed on two occasions over 2 consecutive days, with each interview taking approximately 1 hour. Between interviews, participants wore a wearable camera for continuous, unobtrusive observation, as part of a larger study [15].

Interview questions were semistructured and focused on how community-dwelling people with chronic multimorbidity managed health-related tasks and how they modified their daily lives to accommodate their health needs. The interview question guide is provided in Multimedia Appendix 1 and was developed after consultation with the researchers. The interview questions and procedures were pilot tested with 2 participants before participant recruitment.

All participants were interviewed in their homes based on their choice. Many had arranged for family members to be present, with 38% (10/26) of participants having their spouses present, 4% (1/26) having their child present, 4% (1/26) having their mother present, and 4% (1/26) having their grandchild present. Family members often did not get involved and listened to the conversation. However, they were able to provide information when participants themselves were uncertain and asked for help.

All interviews were conducted by 2 researchers, one of whom was always KY (female), JJ (male), or AYSL (female). All interviews were audio recorded. Field notes were made by the interviewers during the interview to record nonverbal cues and observations inside the dwelling, and notes from both interviewers were consolidated within 24 hours of each interview. The second interview for each patient was conducted by the same researcher pair as in the first interview. The field notes were read alongside interview transcriptions, and any extra information was added to the end of the transcriptions.

Qualitative Data Analysis

Audio recordings were transcribed externally and imported into NVivo Plus (QSR International, version 12). A thematic analysis was conducted inductively-deductively on all interview transcripts.

KY and JJ coded all transcripts separately and established codes that emerged from the data, which described either the work conducted to manage health or how the experience of self-management changed over time. These emerged codes were then deductively placed into themes that aligned with the phases and trajectories in the Chronic Illness Trajectory Model [5] and patient work tasks established in our previous scoping review [16]. Codes that did not fit into any existing themes in either of the frameworks were then read and placed into emergent themes over multiple readings. These new themes were presented alongside themes derived from the two frameworks.

Interview data were analyzed immediately after each interview to detect data saturation, and recruitment ceased when data saturation was reached (defined when no new patient work tasks were being described by 2 participants in a row). KY and JJ reviewed the codes and the theme framework over 6 months, and any discrepancies were resolved via consensus, with monthly meetings. KY, JJ, and AYSL then conducted design ideation [17-19], with brainstorming and scanning the available literature to consider the possible design implications of each need.

In this paper, we report our findings on patient work tasks and how they fit into the Chronic Illness Trajectory Model. Contextual factors influencing tasks and trajectories are beyond the scope of this study.

Results

Participant Demographics

The 26 participants of this study resided across metropolitan Sydney and had a variety of cultural backgrounds (Table 1). The median and mean age were both 72 years (range 46-86), with 61% (16/26) being male and 53% (14/26) identifying as Anglo Australian. The median and mean number of years diagnosed with type 2 diabetes were 19.5 (range 3-50) years. Less than two-third (16/26, 61%) of the participants were using insulin at the time of the study, and 69% (18/26) were retirees. The most common comorbidities were cardiovascular diseases, dyslipidemia, and kidney conditions, with a mean number of 3.96 comorbidities (range 1-20) per person.

XSL•FO
RenderX

Table 1. Participant demographics data (N=26).

Patient demographics	Participant, n (%)
Gender	
Female	10 (38)
Male	16 (62)
Ethnicity	
Anglo Australian	14 (54)
Chinese	4 (15)
Indian	2 (8)
Italian	2 (8)
Trinidad and Tobago	1 (4)
UK migrant	1 (4)
Indonesian	1 (4)
Sri Lankan	1 (4)
Age (years)	
<60	2 (8)
60-64	3 (12)
65-69	3 (12)
70-74	6 (23)
75-79	7 (27)
80-84	2 (8)
85-89	3 (12)
Using insulin	
Yes	16 (62)
No	10 (38)
Major comorbidity (self-identified)	
Cardiovascular conditions	12 (46)
Dyslipidemia	3 (12)
Kidney conditions	3 (12)
Ocular conditions	2 (8)
Thyroid conditions	2 (8)
Prostate conditions	1 (4)
Mental health conditions	1 (4)
Osteoporosis	1 (4)
Traumatic injury	1 (4)
Duration of illness (years)	
<10	3 (12)
10-14	5 (19)
15-19	5 (19)
20-24	5 (19)
25-29	3 (12)
>29	5 (19)
Number of comorbidities	
1	4 (15)

XSL•FO
RenderX

Patient demographics	Participant, n (%)
2	9 (35)
3	1 (4)
4	4 (15)
5	3 (12)
6-10	4 (15)
>10	1 (4)
Employment	
Retired	18 (69)
Self-employed	3 (12)
Employed by others	5 (19)

Phases in the Chronic Illness Trajectory

Table 2 outlines the nine phases in the Chronic Illness Trajectory Model, their definitions [16], and example quotes from our cohort. Table 3 lists the patient work tasks involved in each phase. From our participants, we identified examples of patient work tasks in the following phases: trajectory onset, stable, unstable, acute, crisis, and comeback. More quotes supporting each of the phases can be found in Multimedia Appendix 2. Details of patient work tasks can be found in our previous review [16], which outlines the different task categories, how the categories were defined and created, and what examples were available for each category.

The pretrajectory phase represents the presymptomatic period before symptom presentation. We did not include this phase in our analysis or reporting.

Table 2. Phases within the chronic illness trajectory with definitions and examples.

Phase	Definition	Themes	Examples quotes
Pretrajectory	Before symptom presentation	N/A[a]	N/A
Trajectory onset	Initial symptom presentation and diagnosis	Participants respond to the new diagnosis by contacting health professionals and receiving new information.	“[The endocrinologists] give us a list of what to eat and what not to eat. But sometimes you do it, sometimes you don’t.” [P13, female, age 78 years]
Stable	Symptoms are under control and life activities continue within the limitations of the symptoms	Participants try to overcome inertia and find a new normal to discover what works for them.	“There’s a group online, about 200 people that have all done low-carb [diet], lost 100 pounds...and got their blood A1Cs right down. It seems to be the answer to me.” [P14, male, age 63 years]
Unstable	Symptoms start to get out of control and life activities are adjusted to cope with increasing health demands	Participants react to instability, taking up new tasks, new tools, and new information.	“I’m probably on about 14 [medications] at the moment, because I’ve just had to add two tablets too...when I had my bloods done for my endocrinologist, it came back and I’m very low on iron...he’s put me on iron tablets.” [P6, female, age 72 years]
Acute	Severe exacerbations of symptoms that require normal life activities to be paused	Participants rely on others to maintain basic functionality by prioritizing certain health needs over others.	“When I got told I’m going to be on dialysis, well I had a lot of trouble trying to accept that and kept avoiding it, until I was so sick I had to go on it.” [P11, male, age 76 years]
Crisis	A critical or life-threatening situation where urgent medical care is required	Participants cannot conduct self-management and can only react to crisis points.	“I was not allowed to eat anything. I was not allowed to even drink water, because there was a possibility for surgery at that time.” [P1, male age 67 years]
Comeback	Gradually return to an acceptable level of everyday life	Participants adopt to long-lasting changes and deal with mental distress during adjustment to a new normal.	“Just getting you out of bed and walking, just walking up the end of the corridor and back and that used to exhaust me. But once it’s all over and done with you feel fine. Two weeks of rehab.” [P17, male, age 70 years]
Downward	Consistent decline in health	N/A	N/A
Dying	Final days before death	N/A	N/A

[a]N/A: not applicable.

XSL•FO
RenderX

Table 3. Patient work tasks involved in each phase.

Tasks	Phase					
	Trajectory onset	Stable	Unstable	Acute	Crisis	Comeback
Planning		✓[a]	✓			
Proactive management of risks		✓	✓			✓
Deliberate distraction		✓				
Adapt to social values and expectations		✓				✓
Creating mental coping strategies	✓	✓	✓			✓
Learning about the disease	✓	✓	✓	✓		
Diet control	✓	✓	✓			
Taking treatment	✓	✓	✓	✓	✓	✓
Conduct exercise		✓				✓
Monitor signs and symptoms	✓	✓	✓			✓
Medication management		✓	✓			✓
Self-manage comorbidities		✓	✓			
Use and maintain assistive devices		✓	✓			✓
Do-it-yourself symptom management tools		✓	✓			
Alter the physical environment		✓				✓
Seek medical help	✓	✓	✓	✓	✓	✓
Ask for help from family and friends		✓	✓	✓	✓	✓
Hire professional help		✓	✓			✓
Consult complementary therapy		✓	✓			✓
Search for and attend patient support groups	✓	✓	✓			
Teach others about their health		✓				

[a]The patient work task is involved.

The trajectory onset phase represented when participants first became symptomatic and entered the health system. This phase primarily focused on patients with a gradual disease onset in the Chronic Illness Trajectory Model and was associated with tasks such as visiting doctors and trying to understand medical information. In the stable phase, participants spent the longest time and conducted the most patient work, with a large variety of tasks identified according to our previous publication [16] such as diet control, monitoring signs and symptoms, attending patient support groups, and planning for a new routine.

As health conditions worsen, participants may enter and exit the unstable phase repeatedly, with the goal of patient work being returning to the stable phase. Participants responded to changing health demands in the unstable phase, visiting medical professionals more frequently and starting to use support devices such as walking canes.

The acute and crisis phases represented severe and life-threatening illnesses, respectively. The participants were typically hospitalized during this period. They conducted no patient work, and health professionals oversaw their well-being. In the comeback phase, where recovery and rehabilitation occurred, patient work focused on transitioning back to the stable phase, with specific tasks such as rehabilitating exercises or coping with mental trauma. Under most circumstances, participants would not return to their previous levels of health after an acute, a crisis, or a comeback cycle. Some participants may experience this cycle repeatedly in their lives.

The final two stages of the model, the downward phase (an irreversible deterioration in health) and the dying phase (the last few days before death) were not observed in this study.

Different Types of Chronic Illness Trajectories

Overview

Participants experienced different patterns of change in their illnesses. Some remained stable for most of their disease trajectories, whereas others experienced many crisis episodes. We identified four unique trajectories in our cohort (Figure 2). Each trajectory represented a different life experience and required a different style of patient work adaptation. Each trajectory was also associated with unique goals and needs (Table 4). Multimedia Appendix 3 includes quotes supporting each trajectory. As participants are not at the end of the illness trajectory, it is possible that each participant could experience more than one trajectory in their lifetime. However, given that we are only able to capture past data and cannot predict future events, only the trajectory that each participant was experiencing at the time of the interview was reported here.

XSL•FO
RenderX

Figure 2. Visual representation of the four trajectories.

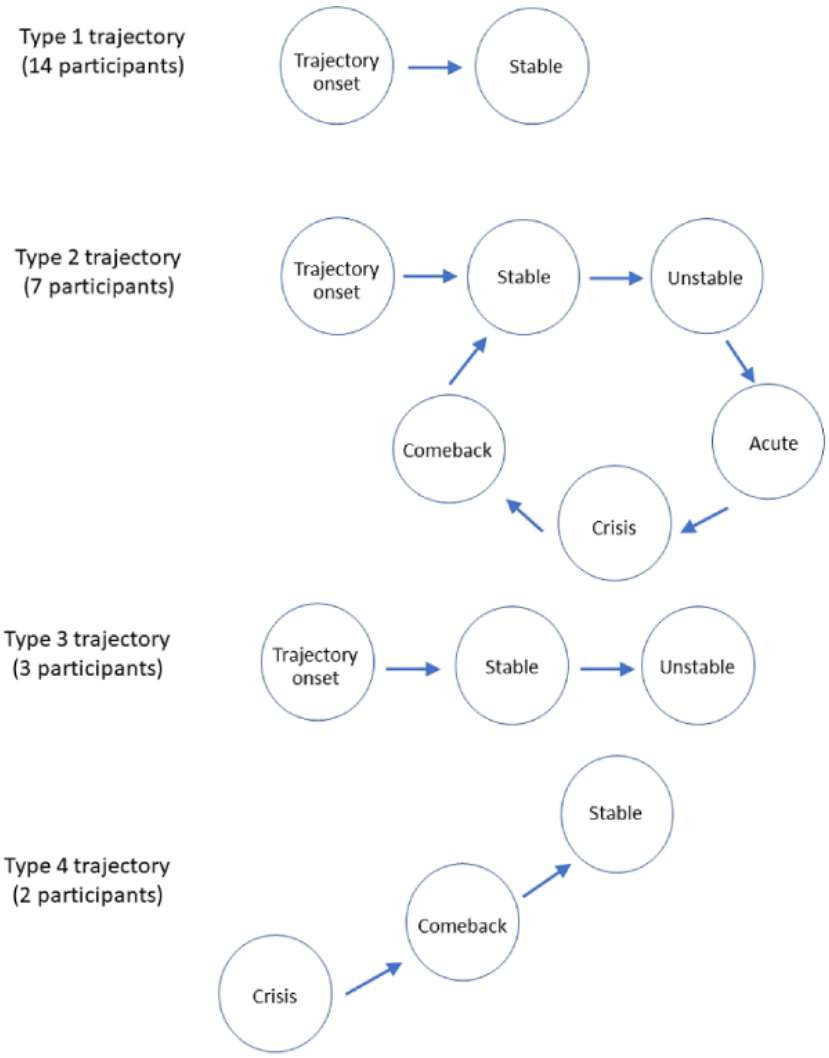


Table 4. Trajectory types and their characteristics (N=26).

Trajectory type	Explanation	Work goals	Participant IDs	Number of comorbidities, mean (SD)
Living with stable chronic conditions	Trajectory onset→stable (participants were diagnosed at a mild stage of the disease where conditions were stable and did not experience disease exacerbations).	Making patient work as routinized and invisible as possible.	7, 9, 10, 14, 15, 19, 20, 22, 23, 25, and 26	3.2 (1.7)
Dealing with cycles of acute or crisis episodes	Trajectory onset→stable→unstable→acute→crisis→comeback→stable (participants have experienced episodes of disease exacerbation, sometimes repeatedly).	Heavily multimorbid, experiencing polypharmacy, and requiring support with self-management adherence.	2, 3, 4, 8, 11, 12, 13, 17, 18, and 21	3.4 (1.7)
Responding to unstable changes in their conditions	Trajectory onset→stable→unstable (participants were diagnosed at a mild stage but are currently experiencing a decline in health).	Experiencing increasing intensity and variety of patient work and dealing with rapid changes.	6, 16, and 24	8.7 (9.7)
Coming back from crisis before stabilizing	Crisis→comeback→stable (participants were diagnosed due to a sudden and severe exacerbation and recovered from that crisis).	Cope with a total loss of normal life and needing holistic support.	1 and 5	1.5 (0.7)

XSL•FO
RenderX

Type 1: Living With Stable Chronic Conditions

The trajectory type most common in our cohort was for those participants who experienced many years of stability with their disease (14/26, 54% of participants). These participants had only experienced the trajectory onset and stable phases (Figure 2) without disease exacerbation. Their work goal was to maximize the integration of patient work into daily life, *making patient work as routinized and invisible as possible*.

Participants in this trajectory type were diagnosed at a mild stage of the disease. They had sufficient opportunity to experiment with patient work tasks, having tried different food options, exercise routines, medications, or information sources, for example. As this group had spent many years in the stable phase, these participants regarded patient work as an incorporated part of their lives. Some participants reported that their patient work had become so ingrained, they no longer remembered each medication's purpose or the roles of the health practitioner they were seeing. Participants averaged 3.2 (SD 3.8) comorbidities in this trajectory, indicating a relatively mild stage of health deterioration.

Type 2: Dealing With Cycles of Acute or Crisis Episodes

The second most common trajectory type was observed in participants who experienced at least one incidence of acute or critical exacerbation followed by the comeback phase (7/26, 27% of participants). Apart from the phases experienced by participants in the type 1 trajectory, the type 2 trajectory also included experiences in the unstable, acute, crisis, and comeback phases (Figure 2). The work goal for this group revolved around being *heavily multimorbid, experiencing polypharmacy, and requiring support with self-management adherence.*

This type of trajectory described participants who had major critical episodes, such as strokes or heart attacks, at least once. Some participants experienced multiple critical episodes and accumulated more medications and diagnoses. Interestingly, due to their complicated medical histories, many participants in this trajectory type had an excellent understanding of medical information. They were aware of the biochemical principles behind their disease and medications but required managerial support to follow their complicated patient work routines. Participants averaged 3.4 (SD 1.7) comorbidities in this trajectory, suggesting a similar level of well-being to the type 1 trajectory group.

Type 3: Responding to Unstable Changes in Their Conditions

The third trajectory type included participants currently in the unstable phase (3/26, 12% of participants). These participants would eventually either proceed onto the acute, crisis, or comeback cycle, or recover enough to return to the stable phase. However, at the time when the interviews were conducted, this group experienced fluctuating health states and had specific work goals and user needs (Figure 2). Their patient work goals showed *increasing intensity and variety of patient work and dealing with rapid changes.*

Owing to their rapidly changing health, participants in this trajectory were seeing their doctors nonroutinely and frequently, with many changes being made to their treatment regimen in a short timeframe. These participants reported confusion regarding the purpose of their treatment and struggled to keep up with their health, often having to rapidly adopt assistive devices (eg, walking canes) or external helpers (eg, hired cleaners). These patients felt that their health was taking over all other spheres of life and required help with understanding how their health was progressing. Participants averaged 8.7 comorbidities (SD 9.7) in this trajectory, significantly higher than the number of comorbidities experienced by the previous two types and suggesting a worse stage of health.

As these participants are currently experiencing exacerbation in health, it is impossible to determine which phase and trajectory they will end up in. For most participants experiencing this trajectory, it is the first time they experience exacerbation, thus lacking the experience and familiarity that participants from the type 2 trajectory may have derived from previous experiences with worsening health. Therefore, we made a distinction for this group of participants, acknowledging the difficulties and stress associated with responding to unstable changes in their health.

Type 4: Coming Back From Crisis Before Stabilizing

The least common trajectory type in our cohort was the participants who received their diagnosis during a crisis event (2/26, 8% of participants). Unlike the other three trajectories, this group initiated their trajectory during the crisis phase (Figure 2), such as with a heart attack or a traffic accident. Their patient work goals focused on *coping with a total loss of normal life and needing holistic support.*

During the crisis phase, the participants were hospitalized for prolonged periods with little to no autonomy. As participants slowly recovered in the comeback phase, they reported experiencing great psychological trauma as they came to understand the irreparable changes in their health. Some participants indicated that they were incapable of coping with these changes during the comeback phase and experienced a range of negative emotions such as dread, devastation, powerlessness, or denial. Suicidal ideation, depression, and thoughts about death were explicitly mentioned. Participants averaged 1.5 comorbidities (SD 0.7) in this trajectory, the lowest number of all types, due to their recent diagnosis and being situated at a relatively earlier stage of the journey of illness despite having experienced significant trauma.

This category of participants did not strictly adhere to the chronic illness trajectory, as participants were not diagnosed at a mild stage of the disease. They also differed from the type 2 trajectory as they did not experience living with their condition in a controlled manner before symptoms worsened and began to go out of control. They do not have the knowledge taught to them by clinicians at diagnosis during the trajectory onset phase and had to learn about self-management while coping with trauma and recovery. Therefore, they have been given a distinct category to reflect their lack of familiarity with this newly diagnosed condition and the few resources compared with those of the other three trajectories previously discussed.

XSL•FO
RenderX

Discussion

Principal Findings

Determining the support needs of multimorbid patients self-managing in the community is an important initial step in designing digital interventions for them. Chronic comorbidities are extremely common in the type 2 diabetes population, with studies indicating that up to 97.5% of patients have one comorbid condition and 88.5% have at least two [20,21]. Our participants were also predominantly affected by comorbidities considered to be concordant with the pathophysiology of type 2 diabetes [22], such as cardiovascular diseases, dyslipidemia, kidney conditions, and ocular conditions. This group represents a population in the community burdened with significant patient work and poses major financial challenges to the health care system if not well-managed. Currently, health apps for chronic patients in the community either target broad populations, such as patients with heart failure, or narrowly defined subgroups, such as patients recently discharged following total knee replacement [23]. Our findings indicate chronic multimorbid patients, such as those with type 2 diabetes, have support needs that evolve over time and are much more complex than those currently supported. As such, a generic app for a certain disease cannot realistically support the differing needs of all patients irrespective of the phase or trajectory type they are in.

Our data revealed four distinct trajectory types over time that produced different self-management goals and work goals, further dissecting chronic multimorbid patients into subgroups based on their previous medical experience. Designing for patient needs from a viewpoint that includes previous medical history and current state is therefore likely to improve user acceptability and appropriateness of digital health apps, enabling such digital tools to provide more timely, suitable, and actionable advice.

Designing for Phase-Specific Needs

Our study indicates that the range and intensity of patient work varies at different phases of the Chronic Illness Trajectory Model. Digital interventions that seek to optimize self-management should therefore be designed according to the specific needs of each phase and acknowledge that different types of support are required for different tasks in each phase.

Studies have begun to address this, moving away from generic *all-patient* user groups, with recent studies designing specific digital tools for newly diagnosed people with diabetes [24] and for people recovering from trauma [25], for example. Such studies have already uncovered significant differences in the subpopulations [11]. Patients using apps also report that as their self-management behaviors change, they can *outgrow* the apps that helped them early in their illness trajectory [26], resulting in continuously seeking out new apps. In our data, patients described different tasks and needs at various phases of the chronic illness trajectory. In the trajectory onset phase, patients commonly thought that the medical information provided was too generic, too technical, and did not translate into actionable suggestions in their own lives. This can be supported by giving personally relevant, precise, and clear advice (such as *eat less bread* and *walk up and down the stairs during lunch break* instead of *eat less carbohydrates* or *do more exercise*). During the unstable phase, patients were compelled to take risk management measures in aspects of life previously taken for granted, such as taping carpets to the floor to prevent slipping or putting rubber bands around stair banisters to feel the stairs at night. Patients in this phase need more information about what is happening to them and what they need to look out for on a daily basis, provided in lay language. The acute and crisis phases produced significant restrictions and burden to patients and their families, with participants describing having to conduct all activities on the mandate of doctors and feeling *shattered* or having *a lot of trouble trying to accept it*. Patients in these phases need mental health support and clear and actionable advice to reduce any chance of confusion or mismanagement. The comeback phase was described as a new lifestyle, with major adjustments to life needed to accommodate changes in routines, such as rehabilitation schedules or dialysis. Patients need logistical support at this stage, such as finding out how to obtain a wheelchair or arrange for subsidized transport, to cope with their reduced health state.

Designing for Trajectories

The four types of trajectories identified in this study correspond to four distinctive design patterns. Each design needs to be tailored to the intended user's needs, digital literacy, environment, and whether the user is the patient or their caregiver. On the basis of the work goals identified in Table 4 and Multimedia Appendix 3, we present the digital needs of the four trajectories in Table 5, together with potential tools that can address their needs. Patients may also shift from one trajectory type to another as their health changes over time, with their need for digital technology changing accordingly.

XSL•FO
RenderX

Table 5. Digital user needs and recommendations for each trajectory type.

Trajectory types	User needs	Potential digital tools
Type 1: living with stable chronic conditions	Tools that normalize patient work and remove the burden of having to think about the disease	• Background data collection tools that require no user input (eg, step-counting phone apps) [27] • Integrated, predetermined lifestyle changes and health intervention delivered automatically (eg, smart fridges that order specific groceries based on existing algorithms and automated prescription refill and delivery) [28]
Type 2: dealing with cycles of acute or crisis episodes	Tools that support self-management adherence and monitors health	• Medication adherence support tools (eg, context-aware digital reminders that cue for medication taking immediately before meals) • Crisis prevention technologies (eg, health monitoring tools that use predictive algorithms to observe signs, such as food and medication consumption, and generate alarms based on behavioral changes)
Type 3: responding to unstable changes in their conditions	Tools that provide symptom monitoring and give alarms for health deterioration	• Crisis prevention technologies (eg, health monitoring tools that use predictive algorithms to observe signs, such as changes in physical symptoms and emotional states, and generate alarms based on symptom changes) • Scheduling and communication assistance (eg, apps that can manage a complicated and changing timetable involving multiple clinicians)
Type 4: coming back from crisis before stabilizing	Tools that support coping with a total loss of normal life and guide patients toward appropriate services and support infrastructure to reestablish normalcy	• Guide the patient to seek appropriate social services (eg, direct patients to appropriate social, financial, and legal services) • Provide support with mental health and coping (eg, phone-based mental health support apps)

Designing for Phase Change

Digital tools that aim to be used throughout the duration of a disease's chronic illness trajectory will need to detect phase changes, such as when the patient's health worsens from the stable to the unstable phase. Although collaborative, co-design exercises with patients to gain insight into user needs are now common during digital health app development, the circumstances of the patients' health do not remain immutable after the app's release. Tools to capture patient-reported experience measure and patient-reported outcome measure [29,30] can collect self-reported data at preset points of the day or immediately after predefined trigger events and can assist with keeping up to date with the patient's health after the app's release. Ideally, an app that detects phase changes should alter its functionality accordingly by activating submodules. For example, participants interviewed for an app designed for mental health expressed a desire for the app to send them mood-regulating messages at times when they were about to lose their temper [31]. This can be achieved either through purely automatic detection of worsening biophysical signs (such as constantly elevated blood pressure or lopsided gait) via external sensors, regular self-reporting by the patient, wearable smart household items such as smart mattresses or smart watches or using individual user data as baselines to train algorithms. Although such data can be entered into digital devices by the patient, automated data collection would reduce the health-related burden of self-monitoring. Excessive requirements for self-reported data could make digital interventions burdensome and contribute to disengagement and dropout over time. Other innovative, context-sensitive digital health interventions, such as *Smart Pill Bottles* that can detect irregularities in medication consumption [32,33], also have the potential to be integrated into a home system that generates external data complementing phone-based apps.

When the interviews took place, each participant's trajectory described the person's current state and journey to this point from the onset of their health conditions, with the future states of each person not necessarily known. Should future research discover these phases and trajectory types to be predictable based on medical history, digital tools can be designed to anticipate such changes and variations between different patients. Individualized and customizable apps [34,35] allow for further tailoring to fit specific subpopulations, and designs of self-care tools would need to adapt to the patient's evolving digital needs to ensure relevance and integration into patient work.

For full realization and evaluation of any of the design implications suggested in this section, co-design sessions would have to be conducted with participants living in the targeted chronic condition, which is beyond the scope of this paper. However, the design implications highlighted here can serve to trigger discussion toward more innovative, holistic, and responsive digital intervention design during co-design sessions, particularly as very few digital interventions currently adapt to changing needs and trajectories.

Limitations

This study was limited in terms of data collection. First, we only recruited participants with a clinical diagnosis who were not severely ill and asked them to recall patient work from the time of diagnosis rather than recruiting individuals at different phases of the chronic illness trajectory. This was done to understand how patient work has evolved for each person. Second, our sample included more males than females because more females declined participation due to family concerns or

XSL•FO
RenderX

obligations. Finally, our recruitment criteria stated that participants must be fluent in English. Consequently, the patient work of people who were not fluent in English was not captured.

Conclusions

This study provides insights into how patient work among multimorbid patients with type 2 diabetes changes over time. There are still gaps in our understanding of how patient health care goals change through different phases of their health, how different patients have different disease trajectories, and how digital health apps can adjust to such changes over time. This study presents data on different types of trajectories, with the perspective of how to use such findings to design better consumer-facing digital health apps. Our findings revealed four different types of trajectories, resulting in different patient work goals. Patients who had never experienced disease exacerbation desired for patient work to be as invisible as possible, whereas those who lived through cycles of crises and recovery needed assistance with self-management adherence. Participants currently experiencing a decline in health needed timely support and crisis prevention technology, and those diagnosed during severe crisis needed guidance to find sources of support and coping. This study highlights opportunities for health informatics and design communities to explore the untapped space of designing for time and trajectory, where future research should incorporate an individual's evolving health experiences when designing digital technologies for patient work over time.

Acknowledgments

The authors would like to thank Mr Wen Jun Liang, Mr Joshua Haddrill, Ms Isabelle Barnes, Mr Christian Tran, Mr Jiawei He, Ms Navy Sea, Ms Emile Chien-Ai Chou, and Ms Natasha Hughes for working as research assistants and acting as chaperones during the interviews. The authors would also like to thank Professor Bernard Tuch, Dr Priscilla Wong, Dr Veronica Prada, and Mr Ivan Khizhnyak for their invaluable help with recruitment. The screenshot conversion software was written by Mr Wen Jun Liang and Mr Joshua Haddrill, without which this study would not have been technically possible. The authors would like to thank Ms Melissa Park for editing earlier versions of the manuscript. Most importantly, the authors would like to thank their participants for their indispensable assistance. They thank the participants for their generosity and kindness, for sharing their stories with us, and for allowing them to record a glimpse of their lives. AYSL was supported by the New South Wales Health Early-to-Mid Career Fellowship, and her research was supported by the National Health and Medical Research Council (NHMRC) Centre for Research Excellence in Digital Health (APP1134919) and NHMRC Centre of Research Excellence in Digital Technology to Transform Chronic Disease Outcomes (ID 1170937). This research was supported by the NHMRC Centre for Research Excellence in Digital Health (APP1134919).

Authors' Contributions

Study conceptualization was carried out by AYSL and KY. Study design was developed by AYSL and KY. Recruitment was carried out by KY, JJ, KWKH, SV, and AYSL. Interviews were conducted by KY, JJ, and AYSL. Data analysis was performed by KY, JJ, AYSL, and EC. The first draft was created by KY, JJ, and AYSL and subsequent drafts by KY, EC, AYSL, KWKH, SV, AB, and FR.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview question guide.
[DOC File , 32 KB - jmir_v23i7e25992_app1.doc]

Multimedia Appendix 2
Quote table for different phases.
[DOC File , 46 KB - jmir_v23i7e25992_app2.doc]

Multimedia Appendix 3
Quote table for different trajectories.
[DOC File , 51 KB - jmir_v23i7e25992_app3.doc]

References

1. Holden RJ, Valdez RS, Schubert CC, Thompson MJ, Hundt AS. Macroergonomic factors in the patient work system: examining the context of patients with chronic illness. Ergonomics 2016 May 10:1-18. [doi: 10.1080/00140139.2016.1168529] [Medline: 27164171]

XSL•FO
RenderX

2. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10. [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
3. Corbin J, Strauss A. Managing chronic illness at home: three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
4. Corbin J, Strauss A. Unending Work and Care: Managing Chronic Illness at Home. 1st Edition. San Francisco: Jossey-Bass; 1988:1-384.
5. Corbin J. The corbin and strauss chronic illness trajectory model: an update. Sch Inq Nurs Pract 1998;12(1):33-41. [Medline: 9805470]
6. Halcomb E, Davidson P. Using the illness trajectory framework to describe recovery from traumatic injury. Contemp Nurse 2005;19(1-2):232-241. [doi: 10.5172/conu.19.1-2.232] [Medline: 16167451]
7. Reed E, Corner J. Defining the illness trajectory of metastatic breast cancer. BMJ Support Palliat Care 2015 Dec;5(4):358-365 [FREE Full text] [doi: 10.1136/bmjspcare-2012-000415] [Medline: 24644176]
8. Burton CR. Re-thinking stroke rehabilitation: the Corbin and Strauss chronic illness trajectory framework. J Adv Nurs 2000 Sep;32(3):595-602. [doi: 10.1046/j.1365-2648.2000.01517.x] [Medline: 11012801]
9. Miller CM. Trajectory and empowerment theory applied to care of patients with multiple sclerosis. J Neurosci Nurs 1993 Dec;25(6):343-348. [doi: 10.1097/01376517-199312000-00004] [Medline: 8106827]
10. Beeney LJ, Bakry AA, Dunn SM. Patient psychological and information needs when the diagnosis is diabetes. Patient Educ Couns 1996 Oct;29(1):109-116. [doi: 10.1016/0738-3991(96)00939-1] [Medline: 9006227]
11. Grobosch S, Kuske S, Linnenkamp U, Ernstmann N, Stephan A, Genz J, GDS Group. What information needs do people with recently diagnosed diabetes mellitus have and what are the associated factors? A cross-sectional study in Germany. BMJ Open 2018 Oct 31;8(10):e017895 [FREE Full text] [doi: 10.1136/bmjopen-2017-017895] [Medline: 30385437]
12. Murphy RP. Management of diabetic retinopathy. Am Fam Physician 1995 Mar;51(4):785-796. [Medline: 7887355]
13. Henricsson M, Nilsson A, Janzon L, Groop L. The effect of glycaemic control and the introduction of insulin therapy on retinopathy in non-insulin-dependent diabetes mellitus. Diabet Med 1997 Feb;14(2):123-131. [doi: 10.1002/(SICI)1096-9136(199702)14:2<123::AID-DIA306>3.0.CO;2-U] [Medline: 9047089]
14. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]
15. Yin K, Harms T, Ho K, Rapport F, Vagholkar S, Laranjo L, et al. Patient work from a context and time use perspective: a mixed-methods study protocol. BMJ Open 2018 Dec 22;8(12):e022163 [FREE Full text] [doi: 10.1136/bmjopen-2018-022163] [Medline: 30580259]
16. Yin K, Jung J, Coiera E, Laranjo L, Blandford A, Khoja A, et al. Patient work and their contexts: scoping review. J Med Internet Res 2020 Jun 02;22(6):e16656 [FREE Full text] [doi: 10.2196/16656] [Medline: 32484449]
17. Kocaballi AB, Ijaz K, Laranjo L, Quiroz JC, Rezazadegan D, Tong HL, et al. Envisioning an artificial intelligence documentation assistant for future primary care consultations: a co-design study with general practitioners. J Am Med Inform Assoc 2020 Nov 01;27(11):1695-1704 [FREE Full text] [doi: 10.1093/jamia/ocaa131] [Medline: 32845984]
18. Hernandez NV, Shah JJ, Smith SM. Understanding design ideation mechanisms through multilevel aligned empirical studies. Des Stud 2010 Jul;31(4):382-410. [doi: 10.1016/j.destud.2010.04.001]
19. Tomitsch M, Wrigley C, Borthwick M, Ahmadpour N, Frawley J, Kocaballi A, et al. Design. Think. Make. Break. Repeat.: A Handbook of Methods. London, United Kingdom: Laurence King Publishing; 2021:1-160.
20. Iglay K, Hannachi H, Howie PJ, Xu J, Li X, Engel SS, et al. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. Curr Med Res Opin 2016 Jul;32(7):1243-1252. [doi: 10.1185/03007995.2016.1168291] [Medline: 26986190]
21. Lang VB, Marković BB. Prevalence of comorbidity in primary care patients with type 2 diabetes and its association with elevated HbA1c: a cross-sectional study in Croatia. Scand J Prim Health Care 2016;34(1):66-72 [FREE Full text] [doi: 10.3109/02813432.2015.1132886] [Medline: 26853192]
22. Piette JD, Kerr EA. The impact of comorbid chronic conditions on diabetes care. Diabetes Care 2006 Mar;29(3):725-731. [doi: 10.2337/diacare.29.03.06.dc05-2078] [Medline: 16505540]
23. Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen W, Hannink G, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
24. Petersen M, Hempler NF. Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):91 [FREE Full text] [doi: 10.1186/s12911-017-0493-6] [Medline: 28651639]
25. Childs BR, Breslin MA, Nguyen MP, Simske NM, Whiting PS, Vasireddy A, et al. Implementation of a mobile app for trauma education: results from a multicenter study. Trauma Surg Acute Care Open 2020;5(1):e000452 [FREE Full text] [doi: 10.1136/tsaco-2020-000452] [Medline: 32548309]

XSL•FO
RenderX

26. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
27. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of personal health records and wearables on health outcomes and patient response: three-arm randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
28. Luo S, Xia H, Gao Y, Jin JS, Athauda R. Smart fridges with multimedia capability for better nutrition and health. In: Proceedings of the International Symposium on Ubiquitous Multimedia Computing. 2008 Presented at: International Symposium on Ubiquitous Multimedia Computing : UMC 2008; October 13-15, 2008; Hobart, Australia p. 39-44. [doi: 10.1109/UMC.2008.17]
29. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]
30. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
31. Goodwin J, Cummins J, Behan L, O'Brien SM. Development of a mental health smartphone app: perspectives of mental health service users. J Ment Health 2016 Oct;25(5):434-440. [doi: 10.3109/09638237.2015.1124392] [Medline: 26732242]
32. Zijp T, Touw D, van Boven J. User acceptability and technical robustness evaluation of a novel smart pill bottle prototype designed to support medication adherence. Patient Prefer Adherence 2020 Mar;Volume 14:625-634. [doi: 10.2147/ppa.s240443]
33. DeMeo D, Morena M. Medication adherence using a smart pill bottle. In: Proceedings of the 11th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT). 2014 Presented at: 11th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT); Oct. 29-30, 2014; Melville, NY, USA. [doi: 10.1109/cewit.2014.7021149]
34. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan;64(1):139-145. [doi: 10.1002/pbc.26177] [Medline: 27574031]
35. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Building a mobile HIV prevention app for men who have sex with men: an iterative and community-driven process. JMIR Public Health Surveill 2015;1(2):e18 [FREE Full text] [doi: 10.2196/publichealth.4449] [Medline: 27227136]

Abbreviations

NHMRC: National Health and Medical Research Council

Edited by R Kukafka; submitted 23.11.20; peer-reviewed by HC Chiu, H Rutjes, M Ozkaynak, A Ritchie, J Hill, K Blondon; comments to author 28.01.21; revised version received 16.02.21; accepted 24.05.21; published 15.07.21.

Please cite as:
Yin K, Jung J, Coiera E, Ho KWK, Vagholkar S, Blandford A, Rapport F, Lau AYS
How Patient Work Changes Over Time for People With Multimorbid Type 2 Diabetes: Qualitative Study
J Med Internet Res 2021;23(7):e25992
URL: https://www.jmir.org/2021/7/e25992
doi:10.2196/25992
PMID:34264193

©Kathleen Yin, Joshua Jung, Enrico Coiera, Kenneth W K Ho, Sanjyot Vagholkar, Ann Blandford, Frances Rapport, Annie Y S Lau. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Human Enough: A Qualitative Study of Client Experience With Internet-Based Access to Pre-exposure Prophylaxis

Shana D Hughes[1], MA, MPH, PhD; Kimberly A Koester[1], PhD; Edvard Engesaeth[2], MD; Merissa V Hawkins[2], MPH; Robert M Grant[1], MPH, MD

[1]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]Nurx, San Francisco, CA, United States

Corresponding Author:
Shana D Hughes, MA, MPH, PhD
Department of Medicine
University of California, San Francisco
550 16th St, 3rd floor
San Francisco, CA, 94158
United States
Phone: 1 415 476 6288
Email: shana.hughes@ucsf.edu

Abstract

Background: HIV pre-exposure prophylaxis (PrEP) is a way to prevent HIV infection using antiretroviral medications. However, common barriers to PrEP engagement include lack of access to prescribers; discomfort seeking sexual health services; and racism, homophobia, and transphobia in medical contexts. Key populations (eg, communities of color, young men who have sex with men, and transgender women) are underrepresented in terms of PrEP uptake in the United States. Nurx is an innovative company that has offered internet-based access to PrEP since 2016.

Objective: In this study, in partnership with Nurx, we aim to explore clients' experiences of digital PrEP access—including the difference made by the telehealth format—and to understand whether Nurx helped reduce barriers to PrEP.

Methods: Electronic chart review and semistructured interviews were conducted with 31 PrEP requesters from California, Florida, Illinois, and New York. Interviews were recorded, transcribed, and subjected to inductive and deductive thematic analysis.

Results: Some interviewees reported initial skepticism about whether a web-based PrEP service could be legitimate or feasible. Despite this, most clients were effusive about their eventual Nurx experience, and many reported that Nurx eased barriers to PrEP access through the availability of knowledgeable, willing prescribers and minimizing embarrassment and discrimination. Our analysis suggests Nurx produced satisfaction by achieving an acceptable balance between 2 client desires: *efficiency* and *humanity*. Efficiency encompasses the simplicity, speed, and convenience of obtaining PrEP, both regarding the Nurx process itself and in comparison with in-person encounters. Humanity covers clients' wish for personalized, responsive interaction and a feeling of connection or care. Nurx's messaging platform was crucial to manifesting these qualities and was largely interpreted through the familiar frame of texting. Clients conceived efficiency and humanity as inversely related in a commercial enterprise and varied in the particular balance they felt was optimal. Those who wished for slightly more humanity than the service afforded used the concept of a *trade-off* to explain why Nurx remained appealing.

Conclusions: Our findings augment evidence that internet-based PrEP provision can broaden access to this HIV prevention strategy. This important finding, notwithstanding a few provisos, merits mention. Telehealth, as practiced by Nurx, was still dependent on culturally competent medical providers as system *inputs*, and the very technology used to overcome access barriers (ie, the internet) generated new hurdles for some clients. Furthermore, clients did not interpret Nurx in a vacuum: their past experiences and the social and structural context mattered. Finally, only granular inquiry revealed precisely *how* Nurx satisfied clients whose experiences and preferences fell within a particular range. Extrapolating from this, we urge scholars not to fetishize technological solutions but rather to interrogate the ways in which any intervention's design works for certain kinds of patients.

(J Med Internet Res 2021;23(7):e22650) doi:10.2196/22650

KEYWORDS

telehealth; internet; HIV; pre-exposure prophylaxis (PrEP); client/user experience; qualitative

XSL•FO
RenderX

Introduction

HIV pre-exposure prophylaxis (PrEP) is a way to prevent HIV infection using antiretroviral medications. The US Food and Drug Administration has approved 2 pharmaceutical products for use as PrEP: tenofovir disoproxil fumarate combined with emtricitabine (for all populations) and tenofovir alafenamide combined with emtricitabine (for adult men and transgender women). PrEP is taken by individuals who are HIV-negative and has been proven to be clinically effective among multiple target populations [1]. Widespread use has already been associated with a significant reduction in new infections in certain local epidemics [2,3]. However, to realize the promise of this biobehavioral HIV prevention strategy, those who are most at risk must opt to take the drugs, and it is well known that some key populations (eg, communities of color, young men who have sex with men, and transgender women) are underrepresented in terms of PrEP uptake in the United States [4,5]. Documented barriers, which may be particularly acute among such populations, include lack of awareness of PrEP, lack of access to or difficulty accessing care, fewer resources to pay for out-of-pocket expenses associated with PrEP (eg, laboratory work), same-sex stigma and racism in medical contexts, and immigration status [6-8].

Telehealth is defined by the US Health Resources and Service Administration as "the use of electronic information and telecommunications technologies to support and promote long-distance clinical health care, patient and professional health-related education, public health and health administration" [9]. This modality of care is not new [10] but is drawing attention as one way to potentially circumvent barriers to PrEP access, especially in the era of COVID-19 when access to specialty care services and care in general is compromised [11]. A recent review [12] noted that PrEP telehealth interventions are being deployed to various ends. Some efforts seek to enhance access among patients who are geographically isolated or encounter challenges locating a knowledgeable provider; others facilitate clinical consultation and/or aim to build capacity among medical providers who may be willing to prescribe PrEP but desire support [5,13-16]. These endeavors are largely the product of partnerships between government, academic, and community stakeholders, and early results suggest that telehealth-based PrEP is feasible and acceptable and may reduce some barriers to access [12], although detailed data on the experiences of patients are scarce.

Private companies have also entered the PrEP telehealth market. One such company in San Francisco, California, called Nurx, offers web- and app-based counseling, prescription, and home-delivery services for PrEP. Nurx began to offer internet-based access to birth control in 2015, expanded to include PrEP a year later, and is now operating in 29 states and the District of Columbia. Qualitative researchers at the University of California, San Francisco (UCSF) partnered with Nurx to explore the possibilities raised by untethering PrEP from traditional brick-and-mortar medical contexts. The initial research objective is to understand how clients make sense of their experience, with special interest in what *difference*, if any, the web-based format makes.

Methods

This research employed a mixed methods design, conducting an electronic chart review and semistructured telephone interviews with clients who requested (although did not necessarily receive) PrEP through Nurx. Sampling so-called *PrEP requesters* was intended to enable the examination of both the experience of completing the envisioned service cycle and the barriers or circumstances that might prompt a prospective Nurx user to initiate but desist from engagement with the service.

Nurx personnel used the platform's asynchronous messaging system to contact all users who had requested PrEP in California. This outreach provided a brief, initial description of study procedures and purpose and requested that users interested in learning more respond with *Yes*. A user's response authorized UCSF interviewers to directly correspond with potential participants through the messaging system and via email (users needed an email address to create a Nurx profile, so this did not limit participation). Participants received a study information sheet via email and asked questions about the research before arranging a telephone interview. Interviewers familiarized themselves with consenting interviewees' cases through chart review. Although we do not directly report data gained through this step in this paper, it was nonetheless important, as it sensitized us to issues that might surface during the interview, such as service gaps, challenges with billing, insurance, or delivery. In some cases, chart review—which included messages exchanged between the clients and the Nurx team—provided a way to triangulate interviewees' accounts. This general procedure was repeated for PrEP requesters in Florida, New York, and Illinois.

The interview guide covered previous knowledge and/or use of PrEP; learning about Nurx; interacting with the web-based platform and Nurx personnel; and receiving, taking, and desisting from PrEP, as relevant. In addition, the guide explored interviewees' motivations for requesting PrEP and sexual practices. Interviews were conducted by KAK and SDH lasted for not more than 90 minutes and were digitally recorded with permission. The interviewees received a US $50 electronic gift card. Data were gathered from April to August 2017, with interviewers jointly debriefing as interviews were conducted. The Institutional Review Board at the UCSF approved all the research procedures.

Interviews were transcribed verbatim, and transcripts were uploaded to MAXQDA 12 Plus [17], a qualitative data analysis software package, to facilitate analysis. Both interviewers participated in the coding process, with transcripts divided evenly between them for first-pass thematic coding. Both deductive and inductive approaches were used to identify salient themes [18]. Deductive codes were drawn from the interview guide (eg, *learned of Nurx* and *adherence*) and the literature (eg, *insurance* and *stigma*). Inductive codes emerged from a close reading of the text itself (eg, *approachable* and *automated messaging*). After the first coding pass, interviewers jointly refined code definitions, and each interviewer reviewed the coding done by her teammate, having the option to confirm the

XSL•FO
RenderX

code applications or modify them in accordance with the revised codebook. Instances in which the revised codebook did not resolve coding discrepancies were examined, and consensus was achieved through discussion. This paper draws on thematic analysis of materials from 31 PrEP requesters in California, New York, Illinois, and Florida. Exemplary quotes in the sections that follow are attributed to interviewees by number to protect their confidentiality.

Results

Sample and Service Flow

Interviewees were predominantly male, aged 30 years or younger, racially and ethnically diverse, highly educated, and identified as gay. Nearly all had insurance, whether public, private, or through parents (Table 1).

Table 1. Participant demographics (N=31).

Characteristic	Participant, n (%)
Sex	
Male	26 (84)
Female	5 (16)
Age (years)	
18-30	25 (81)
31-65	6 (19)
Race or ethnicity	
White (non-Hispanic)	10 (32)
African American (non-Hispanic)	4 (13)
Hispanic or Latino	8 (26)
Asian	3 (10)
Mixed or more multiracial	5 (19)
Insurance	
Uninsured	1 (3)
Medicaid	8 (26)
Parents (unspecified)	5 (19)
Employer based	11 (35)
Kaiser	4 (13)
ACA[a] Plan	2 (6)
Education	
High school or less	3 (10)
Some college, vocational training	9 (29)
Undergraduate degree or more	19 (61)
PrEP[b] status	
Never taken	12[c] (39)
Currently taking	17[d] (55)
Took and stopped	2 (6)
Duration on PrEP (for those who had ever taken PrEP)	
Less than 6 months	9 (47)
More than 6 months but less than 3 years	7 (37)
3 years or more	3 (16)

[a]ACA: Affordable Care Act.

[b]PrEP: pre-exposure prophylaxis.

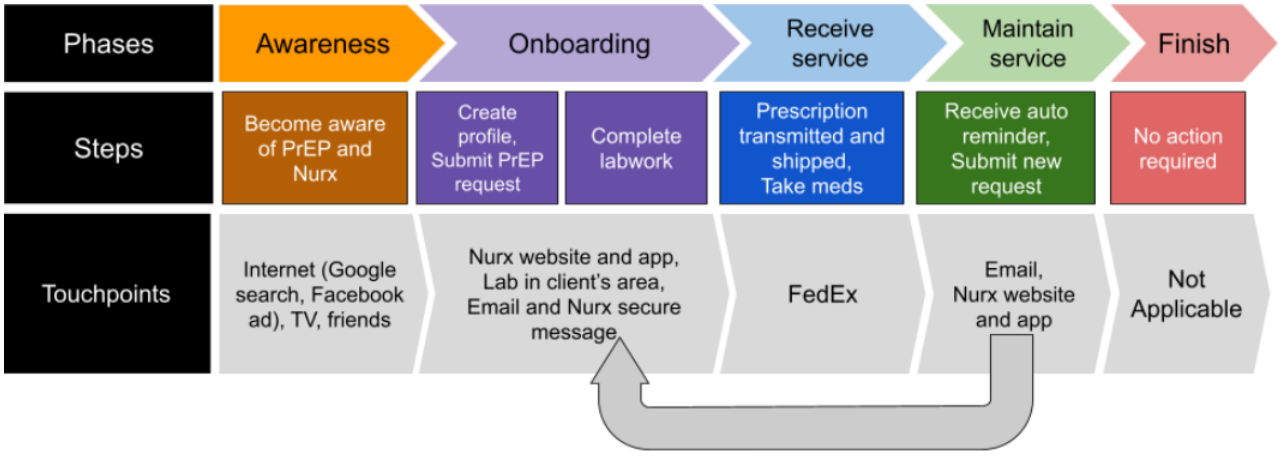
[c]Includes 2 who were awaiting shipments or laboratory orders from Nurx at the time of interview.

[d]Even if no longer receiving pre-exposure prophylaxis from Nurx.

XSL•FO
RenderX

Owing to the novelty of web-based access to PrEP, it is helpful to sketch the Nurx service flow before engaging with the thematic analysis. To do this, we drew from interviewees' descriptions (supplemented with details provided by Nurx staff) to create a simplified journey map of the Nurx client experience, as it was at the time of data collection (Figure 1) [19]. In service design, a client journey is seen to have general phases: awareness, onboarding, receiving the service, maintaining the service, and finally, if necessary, finishing [20].

Figure 1. Simplified Nurx client journey map. PrEP: pre-exposure prophylaxis. TV: television.



In the case of Nurx, clients had to become aware of both PrEP and Nurx. Onboarding included multiple steps. Users created a profile and then initiated a *PrEP request* by answering a web-based questionnaire about sexual behaviors and other HIV-related risks. Those with health insurance also typically provided this information by using a mobile phone to take a photo of their insurance card and uploading that photo to the Nurx site. Personnel at Nurx reviewed this information, and a clinician then sometimes initiated a discussion with the client about the motivation for the PrEP request before sending electronic orders for HIV and other tests (including renal function, hepatitis B and C, sexually transmitted infection [STI], and pregnancy screening, per Centers for Disease Control and Prevention guidance) to a laboratory in the client's area. Laboratories relayed results back to Nurx electronically, and these were reviewed by a Nurx physician.

For clients with nonreactive HIV tests, the first step in the *receiving the service* phase was the transmission of a prescription for PrEP to a pharmacy. However, this happened on *the back end*, meaning it was invisible to clients. They received a message from the physician through the Nurx site that included the results of their laboratory tests, notification that they were being prescribed Truvada, dosing guidance, and other recommendations (eg, PrEP does not protect against STIs other than HIV and so condom use is still recommended). The 3-month prescription was usually shipped directly from the pharmacy to a client-provided address via FedEx. At the time of data collection, the clients had to sign for the package. Once the medication was delivered, the client could begin taking PrEP.

Maintaining the service started with an automated message from Nurx, sent to clients several weeks ahead of prescription renewal time, informing them of the need to submit a new PrEP request through the site. The new PrEP request then generated laboratory orders, and the cycle started again. Clients who did not submit new requests or failed to complete their laboratory tests typically received additional reminders. Prescriptions would be renewed if and when clients' renal function was within the normal range and the HIV test result was negative.

Only 4 participants in this study had experience with finishing. Of 4 participants, 2 men had to stop using the service because of changes in insurance coverage. Another participant stopped using PrEP entirely. The fourth participant might more appropriately be described as *lapsed*, as he had been unable to complete the lab work necessary for his prescription refill. Irrespective of the motivation, finishing did not require the client to do anything; they could simply stop responding to Nurx messages. Even so, several participants used the messaging system to let Nurx know that they would no longer be using the service.

Client Experience: Satisfaction After Overcoming Skepticism

Most interviewees reported being extremely satisfied with Nurx services and peppered descriptions of their experiences with adjectives such as *friendly*, *professional*, and—above all—*convenient*. Illustrative of the general trend, P04 summarized:

> *For someone whose life is busy and unpredictable, this service was like a dream. This really was perfect for my needs.*

Such expressions of satisfaction were common, even among the 10 interviewees who, for various reasons, had not actually accessed PrEP through Nurx or were no longer using the service. As one participant who had moved to a health maintenance organization and found Nurx was no longer covered said:

> *Long story short—any company that is trying to cut [insurance- and cost-related] barriers down, I really root for. So, the only sad thing [about not being able to use Nurx] is that they're going to lose an active user, now. I want their numbers to be up.* [P25]

The pervasive satisfaction was striking because interviewees often related having to overcome an initial hurdle as they

XSL•FO
RenderX

contemplated using Nurx: wondering if the service could possibly be *legit* or whether it was *too good to be true*. This was partially driven by general skepticism about internet content and web-based interactions and was one way in which the web-based format exerted a unique influence on clients' experiences. As such, the telehealth option presents a clear contrast to brick-and-mortar clinical interactions. One interviewee explained:

> *Getting a prescription online, it doesn't seem super legitimate at first...I just had to make sure that...people were using [the website], and...it was working for them before I put my personal information on there...Because I don't want to just hand that to some random person on the Internet.* [P28]

In general, interviewees recounted engaging in *research* intended to ascertain the service's legitimacy. This could include evaluating the Nurx site itself, accessing user reviews, conducting web searches, and reading about the service on trusted news and information sites. In a typical example, one interviewee listed multiple sources that bolstered his confidence in Nurx:

[*The company] had articles talking about it. [The website] looked clean You know, people posted their experiences about it. So, I knew...it's not some scam website that just wanted my health information or something like that*. [P21]

In addition to general concerns about the veracity of web content and the legitimacy of the service, the idea of web-based PrEP access gave some interviewees pause because of what they perceived as potential shortcomings of telehealth. For example, one participant was hesitant because:

> *[Nurx] could very easily not get the whole clinical idea of what's going on with the patient because you're not actually laying eyes on them. And, you know, to me, that just seemed, like, really weird, because I [hadn't] really explored telehealth or anything like that at that point. So, it was just really weird to me. But I was just, like, okay, whatever. I can't get it through my doctor and I really want to get it, so let's reach out through this and try it.* [P24]

PrEP-specific logistical questions also surfaced. Some interviewees who had taken PrEP before learning of Nurx voiced a perception that it might be more complex than telehealth could accommodate. As one participant said:

> *I was like, "This is such a cool idea." There was, I think, not skepticism, more like disbelief because it was so unlike anything that I'd seen before and I couldn't really conceptualize, "Well, how does this really work?" Because this is something that seemed too complicated.* [P04]

Another interviewee, who had been taking PrEP for 2 years before discovering Nurx, said his initial concern had been that:

> *you have to have lab work. So I didn't understand how that would work out [online], until I actually went to their website and read the information.* [P36]

It should be noted that the doubts reported here arise *because of the web-based format*, that is, the very feature that made the service novel and attractive (web-based access and its convenience) also raised barriers to its use. Clients who experienced this (not all did) had to seek out specific information and allay particular suspicions before they felt that PrEP access through telehealth was a truly viable option. These users became sufficiently convinced that Nurx was *legit* (and the system could work) to register and submit requests.

The foregoing notwithstanding, Nurx's service was not flawless. Especially when pressed (eg, "Is there anything at all you would change about the service?"), clients mentioned issues in various parts of the service flow. Some participants had trouble finding the area on Nurx's website that dealt with the PrEP service (P01, P17, P19, and P28), and others explicitly wished for an app (P01, P07, and P21—there might have been more mentions of this had Nurx not rolled out an app during data collection). Some clients had problems at the laboratory they visited (P14, P16, P17, and P25), whereas others encountered challenges with insurance (P02, P15, P20, and P21). Shipping or receiving medication was a frequent source of trouble, especially the FedEx signature requirement (P10, P13, P23, P25, and P36; P02 and P15 also had problems with lost shipments).

To provide more context on these issues, we highlight the case of P07. After signing up with Nurx, he had made a laboratory appointment for the following week and then encountered what he characterized as *a little bit of a rough patch*. This seemed to be a mild description of what was actually one of the most trouble-ridden service journeys in the data set. At his laboratory appointment, a data entry error meant that the staff could not locate the appointment. He made a second appointment but forgot his confirmation number; the laboratory staff said they were unable to help him without it. He made a third appointment a few days later, and that appointment was short and smooth. After subsequently receiving confirmation from Nurx that they could prescribe PrEP for him, he had problems setting up an account with the pharmacy handling the dispensation and delivery. There were issues with his insurance and confusion about how much he would have to pay for the medication. However, at most of these junctures, he mentioned messaging Nurx through the platform and receiving help. He described the challenges he had confronted as "purely, like, circumstantial. It wasn't the process by any means." When asked how the experience of using the service compared with his expectations, he said:

> *It's exceeded actually. I think it's made getting this done...extremely stress-free other than the little hiccup that I had the first time.* [P07]

This is illustrative of one of the strongest trends in the data set: although interviewees could recount challenges with the service, their overall assessment of Nurx was positive. Many participants frequently recommended it to friends and sex partners, and several voiced solidarity with the company. For example, P21 said:

> *I love like the whole startup thing, you know...I want to see Nurx sort of succeed.*

XSL•FO
RenderX

Indeed, only one interviewee's assessment (P24) could reasonably be described as lukewarm overall. Partly because of this, in the rest of this paper, we will pivot to interrogate the pervasive satisfaction that colored the narratives, understanding *how Nurx services produced this* and *for what kinds of clients*, as this can provide a window onto which barriers telehealth actually alleviates.

Client Experience: The Production of Satisfaction

At base, we consider Nurx to have produced satisfaction by achieving an acceptable balance between 2 fundamental client desires, which we have termed *efficiency* and *humanity*. *Efficiency* encompasses the ease of obtaining PrEP, both in terms of how simple and convenient the Nurx process itself is, and the ability to avoid disliked aspects of other ways of getting PrEP. *Humanity* covers the clients' wish for personalized, responsive interaction and a feeling of connection or care. We draw from interviewees' narratives to illustrate these dynamics and their interplay as well as highlight the influence of the web-based format.

> *Efficiency was the way interviewees most clearly contrasted Nurx with traditional clinical contexts. The greatest gains in efficiency accrued when users who otherwise encountered obstacles locating a knowledgeable, willing, available prescriber were able to access PrEP through Nurx [21]. One interviewee, for example, described PrEP as barely a thing where he lived, adding that few health care professionals seemed aware of it. Several of this interviewee's friends had related that:*
>
> *their doctor doesn't know anything about it, and they're not comfortable prescribing it to them. So, they usually get referred out to the [AIDS Health Foundation] in town...So, I was actually going to go there to try to get a prescription, but they were booked out, like, six months or something for appointments.* [P30]

Shortly after learning of the 6-month wait, this interviewee discovered Nurx had expanded its services to his state and he signed up *right away*.

Even users who did not face such extended wait times for appointments described the ability to start the PrEP process at their convenience (ie, by providing initial personal and insurance information through Nurx's web-based portal) as removing a significant constraint. In addition, the quasi-anonymity of submitting answers to behavioral screening questions on the web was highly valued. One participant noted that going to the doctor "for, like, sexual type, gynecologic visits" typically provoked nervousness and stated:

> *I feel that weird shame* [P06]

Another said:

> *If you were speaking to somebody face-to-face about [the] same list of questions that a medical practitioner will ask you [about PrEP], say,..."Do you have unprotected sex with people that may or may not be drug users?" Face-to-face with someone you are not familiar with, that's not a very comfortable question to answer. Now if you get those questions from an online interface, it's a lot easier because there's a potential of not being judged. You're talking to a person, but you're talking to a person via email or vis-à-vis the messaging system. It makes you much more comfortable and much more honest.* [P03]

Nurx's method of obtaining information on patient sexual practices reduced embarrassment, awkwardness, or shame, which these interviewees recounted fearing or having experienced with previous providers. We categorized this as *efficient* because it avoids unnecessary and unpleasant elements of obtaining PrEP in person.

The service flow often went smoothly. When it did not, however, assistance from Nurx—human intervention—was welcomed. For instance, many clients reported great concern over issues of cost and billing and a desire for personalized, real-time responses to questions about these issues. One interviewee reported that his only concern about obtaining PrEP through Nurx had been uncertainty around the cost. After receiving an initial bill of several thousand dollars for a 3-month supply of PrEP, in *sticker shock*, he contacted Nurx and discovered there was a problem:

> *with how my insurance was filing it or something. So, I did have to go back and forth with them a little bit to figure that out. And the price did go down at some point, but it was still over—I think it was, like, $1500 or something was the actual amount that I had to use [the Gilead Patient Assistance Program card] to pay.* [P30]

In such cases, Nurx's ability to respond quickly and individually was key to clients' satisfaction with the service (it bears noting that in this instance, Nurx had proactively signed this user up for the copay card to begin with). With very few exceptions, these interactions occurred through the platform's messaging system. Indeed, when clients talked about the *service* Nurx provided, they mostly discussed messaging. Hence, we examined clients' perceptions of messages in detail.

Messaging

Several interviewees spoke of previously engaging in messaging with a provider through electronic portals that were a part of larger health systems, so this was not always a novel experience. One interviewee reported using such a system to email his primary care provider (PCP), asking to "get [PrEP] started." Once the interviewee learned that he would need to see his PCP in person before starting PrEP, he "decided to try and go through Nurx because [the health system's] availability for check-ups and just anything like that is really impacted" (P08). Although messaging is not necessarily unique to telehealth, what a patient can expect to accomplish by messaging Nurx versus messaging a PCP is likely different.

As previously mentioned, some interviewees noted the potential for feeling less judged through a web-based interface. When asked why this was so, one user (P03, quoted earlier) reflected on providing answers to the sexual behavior questions:

XSL•FO
RenderX

> *Well, people say that emotions or intention is a little bit [harder] to read through text message than it is face-to-face. I honestly feel it kind of applies to what's going on.* [P03]

Interestingly, however, messaging was described as being able to both block the communication of emotion (as mentioned earlier) *and* convey the sense that Nurx was friendly and approachable. As one client explained:

> *The way they write their messages is very cheerful...They addressed you by name like, "Hi, G. my name is so and so and I'm here to collect your insurance information."...I guess I expected it was going to be much more robotic, like, "Please submit such and such information by this date." It wasn't like that at all. It was much more personable, like you were just texting a friend back and forth or something.* [P13]

In this sense, the messaging platform figures in clients' narratives as a tool that Nurx can deploy in the service of both efficiency (*simple and convenient*) and humanity (*just texting a friend*).

Many interviewees described the message system as similar to texting and, thus, familiar. One participant explained why this works in Nurx's favor:

> *The thing that I know is that millennials hate to be inconvenienced. If...they don't have to actually go and talk to someone, I think they will be more inclined to see the process through. Because [also,] the process mimics that of a texting platform that they're used to being on. It mimics that of a social media platform with the way you can post photos to your physician.* [P02]

However, some interviewees directly addressed the use of automated messages, seeing them as distinct from those that felt like *texting a friend*:

> *I can definitely, in the correspondences, tell if it's a form or if it's casual. Obviously, it helps if you ask, like, a bespoke question and receive a bespoke answer. Clearly, that's not coming from an AI bot. But as far as automated, like, "Hey. Thanks so much for filling out your application. We're running at full capacity right now and hope to have your results in the next day or so."...No one sat behind a computer and just casually wrote that before going to the bathroom. You know? That's read over many times by a lot of people, and they thought this was the best they could say.* [P23]
>
> *And how do you feel about that, that there are automated responses going out?* [Interviewer]
>
> *It's a service. It makes sense that it would be automated.* [P23]

Although this interviewee had received both *form* and *casual* responses from Nurx, here, he clearly constructs the feasibility of the service in its web-based format as dependent upon some degree of automation, that is, to remain in business, Nurx as a commercial entity had to serve a sufficient number of clients, which automation made possible. Another interviewee echoed this notion, suggesting that clients sometimes had to accept interactions that felt somewhat *disconnected*. For example, when *specific questions* arose (eg, around insurance or laboratories), messages could be answered by various customer service representatives working the message queue, making him feel:

> *In the flow, it's kind of like, I don't have a specific person to talk to...But, it's an interesting sort of trade-off. Because, at the same time, it was so easy to go through that and...I understand that as being the tradeoff. For like, the ease and simplicity, vs. like, having that like, specific person to contact and like, them being like, readily available.* [P25]

This client's construction of a *trade-off* crystallizes a more implicit notion that was common among client narratives: that efficiency and humanity are inversely related, such that as efficiency increases, humanity decreases and vice versa. The challenge for Nurx was to provide the efficiency clients sought while remaining human *enough*.

A case in point is the experience of actually receiving the medication. Clients described home delivery as an attractive feature of the service, imagined as consummately efficient. However, it was also a touchpoint at which efficiency could falter in practice because of the FedEx signature requirement in effect during data collection (it has since been removed). That a human, physical presence was necessary to complete a transaction touted as virtual was both a conceptual and practical obstacle for many users and detracted somewhat from the seamless way Nurx aimed to fit into its clients' busy, highly mobile lifestyles. Some users had to identify alternative delivery addresses; others consistently had to travel to FedEx locations (or in one case, a local pharmacy itself) to collect their prescriptions. It might be surprising that clients accommodated such demands, given the premium they placed on convenience and efficiency. Our analysis suggests that the concept of a trade-off may help explain. Specifically, it seems that the clients we interviewed were willing to take such actions in exchange for the convenience offered by the other phases of service flow and as long as emergent obstacles throughout the client journey were handled with sufficient humanity. Indeed, it was not only clients who had no other way to obtain PrEP that used the service; those who might otherwise have obtained PrEP from PCPs often chose to remain with Nurx as well. As summarized by one client:

> *Despite all of these complications...getting access to this drug and staying on it—because it has been kind of difficult for me—...each person I've talked to [at Nurx] has, to me, seemed very genuinely concerned and sympathetic and really willing to help me out.* [P15]

This makes plain the importance of the balance that Nurx strikes between efficiency and humanity.

Efficiency-Humanity Balance

The foregoing notwithstanding, within the data set, clients expressed varying levels of tolerance and appreciation for efficiency and humanity. Some gave the impression that they

XSL•FO
RenderX

were close to needing more of the latter than Nurx provided. One interviewee (P01), an ardent proponent of self-monitoring and quantification, had negotiated to have the Nurx physician submit standing orders for STI screening at the laboratory where he did the required testing. This meant that he did not have to wait the customary 3 months; it is unclear how failing to obtain this from Nurx might have impacted his satisfaction with or continued use of the service. Another participant explained that:

> *the problems that I have with [the service] aren't really, like, major issues*

but he recommended Nurx "be more personalized" in their communication with clients, because:

> *using my name isn't going to really cut it. And very rarely do we get into, like, personal conversations unless I initiate something...yes, the anonymity helps, because you just have a little picture in a box [when chatting]. That's great. [But] that could not actually ever be a person. You could be in China, like, cutting or pasting on 27 different things like everybody expects is happening through telehealth.* [P24]

Thus, a lack of *personalization* led the participant to question whether there is *a person* on the other end of the chat at all or whether the person really is who they claim to be. Indeed, never physically seeing the physician seemed to have left a kind of residual uncertainty for several users. One asked:

> *Is [Dr. X, a Nurx physician] a real person?...There's, like, a picture [on the site] of a really pretty lady, and it seems like she might be a stock image or something, you know?* [P30]
>
> *Yeah [...] she's a real person.* [Interviewer]
>
> *Okay. I've been wondering that. Okay. [...]* [P30]
>
> *Was there something else, about the way that she was interacting with you, that made you wonder about that?* [Interviewer]
>
> *No. It was always, like, a really personalized response. So, I knew it wasn't coming from a robot. Yeah. I just kind of—I think it was the photo.* [P30]

This skepticism about the true identity of Nurx service providers echoes the initial doubts some users expressed about the service itself, in that it is engendered by the web-based context of the interaction. However, as with these initial doubts, the discomfort this produced was not sufficient to dissuade these interviewees from using the service.

In contrast to interviewees who seemed to have their needs for *humanity* barely met by Nurx, others appeared to have a much greater tolerance for *efficiency* than the service required, especially if they had existing relationships with other care providers. For these users, a sense of emotional resonance in Nurx's messages or personal connection with the PrEP provider was not that important. For example, in contrasting web-based and face-to-face clinical encounters, one participant said:

> *[Online is] not as personable, you know, but that's honestly not a big deal to me...I like to get in there, get everything taken care of, and go, you know? We don't have to, you know, ham it up and just have this great, friendly relationship.* [P27]

These data suggest that users arrive at the Nurx clinical encounter with a set of individual preferences and needs regarding the patient-provider relationship, an orientation that circumscribes a range within which interactions will be considered acceptable. At one end, some clients seemed to wish for slightly more humanity than was built into the service's default operation; at the other end, clients were satisfied though basically uninterested in having a relationship with Nurx that went beyond the transactional. Similarly, Nurx is configured in ways that enable a range of service-user interactions (those encompassed within the particular efficiency-humanity balance struck by the company) and make others less likely or impossible. When the user and company ranges overlap, clients are likely to judge the resulting interactions as satisfactory, as long as their end goal (eg, obtaining PrEP) is also achieved. On the basis of our data, the ideal Nurx client is a person who has access to and is relatively comfortable using the internet, did not experience or was able to overcome initial skepticism, feels their schedule is busy enough that all kinds of real-time appointments are a chore, may wish to avoid face-to-face discussions of sexual practices, and/or does not require an intensely personalized relationship with the PrEP provider.

Discussion

Nurx: Telehealth Considerations and Lessons

Although telehealth-based PrEP provision is increasingly mentioned as a potential way to broaden access to this HIV prevention strategy [22] and a handful of innovative interventions have appeared in the literature [12], fine-grained data on the patient or user experience are scarce. The partnership between UCSF researchers and Nurx enabled the collection of qualitative data to shed light on this topic from within a commercial environment. Our analysis led us to explain client satisfaction as resulting from the balance the company struck between what we have termed *efficiency* (convenience, ease, and automation) and *humanity* (personalized interaction and care). The concept of a *trade-off* explains how certain highly valued parts of the service (eg, convenient, flexible service initiation; quasi-anonymous discussion of potentially sensitive sexual health topics; and messaging platform and practices) helped users accept other relative inconveniences (eg, delivery issues or, in some cases, presenting for follow-up laboratory testing [23]). In addressing the important questions of whether Nurx provided a way to overcome common barriers to PrEP access and what difference the web-based format made, we ground the discussion firmly in our data while answering with an eye toward telehealth more generally. In light of the massive surge in interest and use of web-based formats for medical appointments prompted by the current global COVID-19 pandemic, we hope that scholars and practitioners will find the insights and questions we raise transferable [24] and valuable.

We found evidence that Nurx allowed patients to locate a knowledgeable and willing prescriber (as long as they lived in a state served by the company). In addition, receiving PrEP services from Nurx enabled users to avoid stigmatizing or

XSL•FO
RenderX

embarrassing encounters with providers. These outcomes are important and should not be minimized, particularly as they manifested among a diverse patient population that included young men of color who have sex with men, a group disproportionately impacted by HIV [25]. That impact notwithstanding, in assessing Nurx as an intervention, we should distinguish between outcomes that are produced by, as opposed to merely facilitated by, telehealth. Connecting providers and patients despite the geographic distance between them and reducing the burden of medical appointments (by removing the need to travel to and from) are inherent capabilities of telehealth. In contrast, that Nurx personnel and physicians were experienced as nonjudgmental during potentially sensitive *conversations* is not *produced by* the web-based format, in the sense of being guaranteed simply by virtue of happening on the web. While technology enabled users to interact with Nurx personnel without being face-to-face (which was said to remove some potential for embarrassment), the content and tenor of those exchanges also certainly influenced users' comfort level. Both Nurx's underlying company philosophy and their experience of providing oral contraception may have contributed to their capacity to create interactions users described as nonjudgmental and sex positive. If the providers on the other end of the messaging platform had lacked cultural competence in providing sexual health services [26], it is unlikely that users would have felt so at ease. This points to the importance of particular types of providers and messaging as components of a telehealth system [21].

As mentioned earlier, although Nurx helped patients who had previously encountered difficulties in trying to access PrEP, interviewees also recounted experiencing substantial skepticism or uncertainty about the service, especially initially. Although some of these doubts were PrEP related (eg, how would lab testing work?), others grew out of the remote or web-based modality of care and may not be specific to Nurx or PrEP. Thus, regarding telehealth generally, it is imperative to recognize that *the web-based format itself may engender barriers to be overcome*. This possibility tends to be absent from discussions of the promise of telehealth. Those wishing to use telehealth approaches (whether in health interventions or as commercial entities) should consider what information or strategies would prevent such doubts from arising in the first place, or at least how to allay them should they arise.

Another crucial point to make, based on stories interviewees shared, was that much of what influenced their assessment of Nurx lies outside of the telehealth experience itself. Especially when confronted with novel situations, users of the service drew on previous experiences to inform their notions about acceptability and appeal. For example, most users were unaccustomed to messaging back and forth in anything close to real time with a medical provider. However, such behavior was familiar from other domains (eg, text messaging and social media posts), and it seemed that not only the practice but also the meanings attached to that practice—that is, friendliness and responsiveness—became associated with Nurx. On the other hand, past *real-life* experiences were sometimes contrasted with what occurred via telehealth. In particular, most clients drew on face-to-face medical appointments as the implicit benchmark against which they evaluated their Nurx experience. As in-person appointments often came with what clients saw as hassles, such as lack of convenient appointment times, the embarrassment of talking about sexual health and practices, and providers who were not knowledgeable about or supportive of PrEP, inconveniences encountered with Nurx seemed minimal in comparison (the *trade-off*). When creating telehealth (including mobile health) services, it may be helpful to explore what previous experiences users could draw on to help them interpret their new telehealth experiences in a positive light and use design to maximize these associations of relative advantage.

In addition, although often not explicit in interviewees' accounts, any social determinants of health perspective [27] will acknowledge that cultural and structural factors play a role in the clients' assessment of Nurx. For instance, the degree to which clients anticipate or have experienced embarrassing or shame-inducing sexual health discussions with providers may depend on the cultural norms of the community where they live or have previously received care, which, in turn, likely undergirds how appealing it would be to avoid such encounters, and therefore how attractive clients find Nurx's model. Examples of structural influences are geographic distribution of PrEP-knowledgeable providers, what kind of insurance clients had (if any), and whether they lived in states that had expanded Medicaid, all of which impact PrEP accessibility. To the extent that national policy influences insurance coverage (eg, both the enactment and dismantling of the Affordable Care Act), it should also be accounted for. Although it is understandable and appropriate that assessments of technology-based health interventions focus on feasibility, acceptability, process metrics, and behavioral and clinical outcomes [28-32], we urge scholars not to leave the wider social context in which telehealth is embedded entirely *outside the frame*.

One further point we wish to raise for discussion is the pressing need for fine-grained inquiry into PrEP telehealth. Our research indicated features of Nurx that were key to the appeal and users' experience of the service, as well as revealed variation in terms of the overall balance interviewees desired between efficiency and humanity. In short, patients exhibited a range of needs and preferences for clinical encounters, and Nurx, like any technological intervention, was built to accommodate a particular range of these preferences. When these ranges overlapped, and the end goal (obtaining PrEP) was achieved, satisfaction was often the result.

A serious limitation of scholarly production on nontraditional forms of PrEP access is that it rarely engages in a meaningful way with the differences in format, service design, or context of implementation among interventions. As such, we agree with Mayer et al [7] that "studies that identify the core components of effective programmatic partnerships are needed." They continue, however, to posit the goal of such work as the development of "normative guidance" and the promotion of "best practices for local PrEP implementation programs" [7]. We acknowledge *best practices* as potentially helpful and effective *as guidelines*, but as we have argued elsewhere [21], we believe that interventions, contexts, and users are mutually constitutive, which is to say that, in different contexts or with different users, even the same telehealth intervention might

XSL•FO
RenderX

mean something different and hence function quite differently [33,34]. Indeed, *best practices* may vary so widely across different locales, key populations, or user orientations (eg, to clinical interactions) so as to preclude a singular *right* answer about *how to do* telehealth. Even within the same context, clients with different needs and preferences than the ones we interviewed could easily evaluate Nurx quite differently. For example, individuals with serious privacy concerns about the web or those who could not get over their worries that Nurx is a scam might be more reassured by and therefore more likely to opt for a service that featured real-time videoconference or telephone appointments with a medical provider. Different flavors of telehealth will appeal to different kinds of patients. Rather than assuming that we know how client-patients experience these interactions, researchers should strive to illuminate the mechanics at play, that is, we need to ask the following questions: What *is* the client's experience? *How* is this experience produced? F*or what kind of clients* does this hold true? Digging into the specifics of the client experience in this way allows us to avoid fetishizing the modality of service delivery—after all, technology is a tool, not an end in itself.

Limitations

As with any research, this study has important limitations, mostly to do with potential selection bias among our interviewees. Most obviously, people who lack access to the internet or are unaware of PrEP and/or Nurx could not have registered as Nurx users. In addition, potential users who felt the need for an intensely *hands-on* experience with a medical provider around PrEP would likely not have found Nurx services attractive in the first place, and those who experienced but were not able to overcome strong skepticism about Nurx probably would not have registered as users on the site. Thus, we were unable to interview anyone representing these groups.

Among the individuals we did interview, several notable categories of potential clients may be underrepresented. First, none of our interviewees identified as American Indian or Alaskan Native, Native Hawaiian and other Pacific Islander, or transgender. The latter is particularly lamentable because people of trans experience figure among those for whom telehealth has been proposed as a way to circumvent barriers to care [35]. Second, although we opted to recruit *PrEP requesters* (rather than only those who had received PrEP through Nurx) in a bid to reach individuals who had negative experiences, those users may have been less likely than their satisfied counterparts to be interested in participating in this research. However, as our analysis suggests that patients with different needs may gravitate to other ways of obtaining PrEP (via a different approach to telehealth or in a brick-and-mortar context), this simply means that other studies should pursue a similar detailed line of inquiry about patient experiences with those other methods. By comparing and contrasting our findings, we may be able to derive key characteristics or needs that would indicate how and for whom differently designed services work best. Finally, we recognize that users who lacked other ways to access PrEP may have been especially likely to report feeling positive about Nurx, even if they experienced challenges in using the service. However, this would be true of *any* PrEP service when a user has no other mode of access, and that other options are unavailable or unappealing is part of what produces satisfaction with real-world clinical settings. While acknowledging this, we do not believe it undermines our analysis, as it is focused primarily on understanding not how many interviewees said they were satisfied but how that satisfaction is produced.

Conclusions

Interviewees recognized the web-based nature of Nurx, with its potential for real-time communication and use of automation, as a unique platform that enabled a novel form of PrEP access. However, there was variation among clients in the efficiency-humanity balance they wanted Nurx to strike. Some clients very heavily valued efficiency and seemed to see Nurx as almost analogous to other commercial entities with whom they might have a subscription (eg, Netflix). Others demanded a more personal touch and seemed to interpret the relationship they had with Nurx—technological mediation notwithstanding—as a caring one. However, for all clients, every step in the Nurx journey was simultaneously technological, clinical, and social, informed by previous experiences in various domains (eg, in-person medical appointments and texting), as well as cultural and structural considerations. While telehealth is not a panacea, having multiple, differently designed access options available may fit the needs of the broadest swath of potential users, thereby opening new spaces for therapeutic engagement.

Acknowledgments

The authors would like to thank the research participants for sharing their experiences, the entire Nurx team for a cheerful and productive collaboration, Dr Chris Hall for valuable feedback on the draft manuscript, and Lauren Sterling at the UCSF-Gladstone Center for AIDS Research and our financial analysts at the Division of Prevention Science for their support. #teamscience

Conflicts of Interest

None declared.

References

1. Fonner V, Dalglish SL, Kennedy CE, Baggaley R, O'Reilly KR, Koechlin FM, et al. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. AIDS 2016 Jul 31;30(12):1973-1983 [FREE Full text] [doi: 10.1097/QAD.0000000000001145] [Medline: 27149090]

XSL•FO
RenderX

2. HIV Epidemiology Annual Report. San Francisco Department of Public Health. 2018. URL: https://r.search.yahoo.com/_ylt=AwrxhSS82YZg8GAA_jrnHgx.;_ylu=Y29sbwMEcG9zAzEEdnRpZAMEc2VjA3Ny/RV=2/RE=1619479101/RO=10/RU=https%3a%2f%2fwww.ecdc.europa.eu%2fen%2finfectious-diseases-public-health%2fhiv-infection-and-aids%2fsurveillance-and-disease-data%2fannual-0/RK=2/RS=dKg9DTH19YdUFnCRDwPmf1p7OXc- [accessed 2018-12-31]
3. HIV in Australia: Annual Surveillance Short Report. Kirby Institute. 2018. URL: https://kirby.unsw.edu.au/report/hiv-viral-hepatitis-and-sexually-transmissible-infections-australia-annual-surveillance [accessed 2021-05-31]
4. Harawa N, McBride S, Leibowitz A, Pulsipher C, Holloway I. Examining PrEP Uptake among Medi-Cal Beneficiaries in California: Differences by Age, Gender, Race/Ethnicity and Geographic Region. California HIV/AIDS Policy Research Centers 2018 Feb:1-9.
5. Refugio O, Kimble MM, Silva CL, Lykens JE, Bannister C, Klausner JD. Brief Report: PrEPTECH: A Telehealth-Based Initiation Program for HIV Pre-exposure Prophylaxis in Young Men of Color Who Have Sex With Men. A Pilot Study of Feasibility. J Acquir Immune Defic Syndr 2019 Jan 01;80(1):40-45 [FREE Full text] [doi: 10.1097/QAI.0000000000001873] [Medline: 30272632]
6. Pinto RM, Berringer KR, Melendez R, Mmeje O. Improving PrEP Implementation Through Multilevel Interventions: A Synthesis of the Literature. AIDS Behav 2018 Jun 5;22(11):3681-3691. [doi: 10.1007/s10461-018-2184-4]
7. Mayer K, Chan PA, Patel R, Flash CA, Krakower DS. H.., Evolving Models and Ongoing Challenges for HIV Preexposure Prophylaxis Implementation in the United States. Journal of acquired immune deficiency syndromes (1999) 2018;77(2):119-127. [doi: 10.1097/qai.0000000000001579]
8. Calabrese SK, Earnshaw VA, Underhill K, Hansen NB, Dovidio JF. The Impact of Patient Race on Clinical Decisions Related to Prescribing HIV Pre-Exposure Prophylaxis (PrEP): Assumptions About Sexual Risk Compensation and Implications for Access. AIDS Behav 2013 Dec 24;18(2):226-240. [doi: 10.1007/s10461-013-0675-x]
9. Telehealth programs. Health Resources and Services Administration (HRSA). URL: https://www.hrsa.gov/rural-health/telehealth [accessed 2020-06-13]
10. Mahar JH, Rosencrance GJ, Rasmussen PA. Telemedicine: Past, present, and future. CCJM 2018 Dec 01;85(12):938-942. [doi: 10.3949/ccjm.85a.17062]
11. AIDS Vaccine Advocacy Coalition (AVAC), PATH, and International AIDS Society (IAS). Bringing PrEP closer to home: Why is now the time for differentiated PrEP? In: Non-commercial satellite session. 2020 Presented at: AIDS 2020; 6 July 2020; Virtual p. 1-1 URL: https://www.aids2020.org/satellite-symposia/#toggle-id-22
12. Touger R, Wood BR. A Review of Telehealth Innovations for HIV Pre-Exposure Prophylaxis (PrEP). Curr HIV/AIDS Rep 2019 Jan 30;16(1):113-119. [doi: 10.1007/s11904-019-00430-z]
13. Hoth AB, Shafer C, Dillon DB, Mayer R, Walton G, Ohl ME. Iowa TelePrEP: A Public-Health-Partnered Telehealth Model for Human Immunodeficiency Virus Preexposure Prophylaxis Delivery in a Rural State. Sexual Trans Dis 2019 May 31;46(8):507-512. [doi: 10.1097/olq.0000000000001017]
14. Stekler J, McMahan V, Ballinger L, Viquez L, Swanson F, Stockton J, et al. HIV Pre-exposure Prophylaxis Prescribing Through Telehealth. JAIDS Journal of Acquired Immune Deficiency Syndromes 2018;77(5):e40-e42.
15. Chartier M, Gylys-Cowell I, Van Epps P, Beste LA, Ohl M, Lowy E, et al. Accessibility and Uptake of Pre-Exposure Prophylaxis for HIV Prevention in the Veterans Health Administration. Fed Pract 2018 Mar;35(Suppl 2):S42-S48. [Medline: 30766393]
16. Wood BR, Mann MS, Martinez-Paz N, Unruh KT, Annese M, Spach DH, et al. Project ECHO: telementoring to educate and support prescribing of HIV pre-exposure prophylaxis by community medical providers. Sex. Health 2018;15(6):601. [doi: 10.1071/sh18062]
17. MAXQDA 12 Plus. VERBI Software. 2015. URL: https://www.maxqda.com/products/maxqda-release-notes# [accessed 2015-12-31]
18. Ryan GW, Bernard HR. Techniques to Identify Themes. Field Methods 2016 Jul 24;15(1):85-109. [doi: 10.1177/1525822x02239569]
19. Sultani L. Customer Journey Map Template. 12 July. 2013. URL: http://lesliedesigns.com/wp/2013/05/23/customer-journey-map-template/ [accessed 2018-07-12]
20. Kalbach J. Mapping Experiences: A Complete Guide to Creating Value through Journeys, Blueprints, and Diagrams. Sebastopol, CA: O'Reilly Media; 2016.
21. Hughes S, Koester K. Internet-Based Access to PrEP in the US: A 'Critically Applied' Approach and the Symbolic Effects of a Clinical-Technological Assemblage. Somatosphere: Science, Medicine, Anthropology. 2019. URL: https://r.search.yahoo.com/_ylt=AwrxgzOK3oZgKhYAvA_nHgx.;_ylu=Y29sbwMEcG9zAzEEdnRpZAMEc2VjA3Ny/RV=2/RE=1619480331/RO=10/RU=http%3a%2f%2fsomatosphere.net%2f2019%2finternet-based-access-to-prep-in-the-us-a-critically-applied-approach-and-the-symbolic-effects-of-a-clinical-technological-assemblage.html%2f/RK=2/RS=DXoZM611hStGDMuEWYwSBAiqLRo- [accessed 2019-07-22]
22. Sullivan PS, Siegler AJ. Getting pre-exposure prophylaxis (PrEP) to the people: opportunities, challenges and emerging models of PrEP implementation. Sex. Health 2018;15(6):522. [doi: 10.1071/sh18103]
23. Koester KA, Hughes SD, Grant RM. “A Good Habit”: Telehealth PrEP Users Find Benefit in Quarterly Monitoring Requirements. J Int Assoc Provid AIDS Care 2020 Apr 23;19:232595822091926. [doi: 10.1177/2325958220919269]

XSL•FO
RenderX

24. Kitto SC, Chesters J, Grbich C. Quality in qualitative research. Medical Journal of Australia 2008 Feb 18;188(4):243-246. [doi: 10.5694/j.1326-5377.2008.tb01595.x]
25. HIV Surveillance Report. Centers for Disease Control and Prevention. 2019. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-07-07]
26. Knight RE, Shoveller JA, Carson AM, Contreras-Whitney JG. Examining clinicians' experiences providing sexual health services for LGBTQ youth: considering social and structural determinants of health in clinical practice. Health Education Research 2014 Jan 10;29(4):662-670. [doi: 10.1093/her/cyt116]
27. Dean HD, Fenton KA. Addressing Social Determinants of Health in the Prevention and Control of HIV/AIDS, Viral Hepatitis, Sexually Transmitted Infections, and Tuberculosis. Public Health Rep 2010 Jul;125(4_suppl):1-5. [doi: 10.1177/00333549101250s401]
28. Feldman J, Milne J, Mateo K, Ahmed N, Odlum M, Pugliese L, et al. Conference Program. 2019 Presented at: American Public Health Association Annual Meeting and Expo; 2-6 November 2019; Philadelphia, PA p. 1-1 URL: https://apha.confex.com/apha/2019/meetingapp.cgi/Paper/445532
29. Koneska E, Appelbe D, Williamson PR, Dodd S. Usage Metrics of Web-Based Interventions Evaluated in Randomized Controlled Trials: Systematic Review. J Med Internet Res 2020 Apr 16;22(4):e15474. [doi: 10.2196/15474]
30. Eysenbach G. CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923]
31. Soriano JB, García-Río F, Vázquez-Espinosa E, Conforto JI, Hernando-Sanz A, López-Yepes L, et al. A multicentre, randomized controlled trial of telehealth for the management of COPD. Respiratory Medicine 2018 Nov;144:74-81. [doi: 10.1016/j.rmed.2018.10.008]
32. Vandekerckhove P, Vandekerckhove Y, Tavernier R, De Jaegher K, de Mul M. Leveraging User Experience to Improve Video Consultations in a Cardiology Practice During the COVID-19 Pandemic: Initial Insights. J Med Internet Res 2020 Jun 25;22(6):e19771. [doi: 10.2196/19771]
33. Rhodes T, Closson EF, Paparini S, Guise A, Strathdee S. Towards “evidence-making intervention” approaches in the social science of implementation science: The making of methadone in East Africa. International Journal of Drug Policy 2016 Apr;30:17-26. [doi: 10.1016/j.drugpo.2016.01.002]
34. Holt M. Configuring the users of new HIV-prevention technologies: the case of HIV pre-exposure prophylaxis. Culture, Health & Sexuality 2014 Oct 23;17(4):428-439. [doi: 10.1080/13691058.2014.960003]
35. Stephenson R, Todd K, Kahle E, Sullivan SP, Miller-Perusse M, Sharma A, et al. Project Moxie: Results of a Feasibility Study of a Telehealth Intervention to Increase HIV Testing Among Binary and Nonbinary Transgender Youth. AIDS Behav 2019 Nov 23;24(5):1517-1530. [doi: 10.1007/s10461-019-02741-z]

Abbreviations

PCP: primary care provider
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 19.07.20; peer-reviewed by N Subedi, P Meylakhs; comments to author 17.11.20; revised version received 16.02.21; accepted 03.04.21; published 05.07.21.

Please cite as:
Hughes SD, Koester KA, Engesaeth E, Hawkins MV, Grant RM
Human Enough: A Qualitative Study of Client Experience With Internet-Based Access to Pre-exposure Prophylaxis
J Med Internet Res 2021;23(7):e22650
URL: https://www.jmir.org/2021/7/e22650/
doi:10.2196/22650
PMID:36256828

©Shana D Hughes, Kimberly A Koester, Edvard Engesaeth, Merissa V Hawkins, Robert M Grant. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Telerehabilitation to Deliver a Home Exercise Program to Youth With Arthrogryposis: Single Cohort Pilot Study

Marianne Gagnon[1,2], MSc; Gabriela Marino Merlo[1], DCS; Rita Yap[1], MSc; Jessica Collins[1], MSc; Caroline Elfassy[1,3], MSc; Bonita Sawatzky[4], PhD; Jacquelyn Marsh[5], PhD; Reggie Hamdy[1,6], MD, FRCSC, MSc; Louis-Nicolas Veilleux[1,2], PhD; Noémi Dahan-Oliel[1,3], PhD

[1]Shriners Hospitals for Children-Canada, Montreal, QC, Canada

[2]Department of Surgery, McGill University, Montreal, QC, Canada

[3]School of Physical and Occupational Therapy, McGill University, Montreal, QC, Canada

[4]Department of Orthopedics, University of British Columbia, Vancouver, BC, Canada

[5]School of Physical Therapy, Western University, London, ON, Canada

[6]Division of Paediatric Orthopaedics, Department of Paediatric Surgery, Montreal Children Hospital, Montreal, QC, Canada

Corresponding Author:
Noémi Dahan-Oliel, PhD
Shriners Hospitals for Children-Canada
1003 Décarie Blvd
Montreal, QC
Canada
Phone: 1 514 842 4464 ext 2278
Email: noemi.dahan@mcgill.ca

Abstract

Background: Arthrogryposis multiplex congenita (AMC) is characterized by joint contractures and muscle weakness, which limit daily activities. Youths with AMC require frequent physical therapeutic follow-ups to limit the recurrence of contractures and maintain range of motion (ROM) and muscle strength; however, access to specialized care may be limited because of geographical distance. Telerehabilitation can offer a potential solution for delivering frequent follow-ups for youth with AMC, but research on the use of telerehabilitation in children with musculoskeletal disorders is scarce.

Objective: The study aims to evaluate the feasibility of delivering a home exercise program (HEP) by using telerehabilitation for youth with AMC. We also aim to explore the effectiveness of the HEP as a secondary aim.

Methods: Youths aged between 8 and 21 years with AMC were recruited at the Shriners Hospitals for Children–Canada. The participants completed baseline and post-HEP questionnaires (the Physical Activity Questionnaire for Adolescents, Pediatrics Outcomes Data Collection Instrument, and Adolescent and Pediatric Pain Tool), and clinicians assessed their active ROM using a virtual goniometer. Clinicians used the Goal Attainment Scale with the participants to identify individualized goals to develop a 12-week HEP and assess the achievement of these goals. Follow-ups were conducted every 3 weeks to adjust the HEP. Data on withdrawal rates and compliance to the HEP and follow-ups were collected to assess the feasibility of this approach. The interrater reliability of using a virtual goniometer was assessed using the intraclass correlation coefficient and associated 95% CI. Nonparametric tests were used to evaluate feasibility and explore the effectiveness of the HEP.

Results: Of the 11 youths who were recruited, 7 (median age: 16.9 years) completed the HEP. Of the 47 appointments scheduled, 5 had to be rescheduled in ≤24 hours. The participants performed their HEP 2.04 times per week (95% CI 1.25-4.08) and reported good satisfaction with the approach. A general intraclass correlation coefficient of 0.985 (95% CI 0.980-0.989) was found for the web-based ROM measurement. Individualized goals were related to pain management; endurance in writing, standing, or walking; sports; and daily activities. In total, 12 of the 15 goals set with the participants were achieved. Statistically significant improvements were observed in the pain and comfort domain of the Pediatrics Outcomes Data Collection Instrument (preintervention: median 71; 95% CI 34-100; postintervention: median 85; 95% CI 49-100; *P*=.08) and Physical Activity Questionnaire for Adolescents (preintervention: median 1.62; 95% CI 1.00-2.82; postintervention: median 2.32; 95% CI 1.00-3.45; *P*=.046).

XSL•FO
RenderX

Conclusions: The remote delivery of an HEP for youth with AMC is feasible. Promising results were found for the effectiveness of the HEP in helping youths with AMC to achieve their goals. The next step will be to assess the effectiveness of this exercise intervention in a randomized controlled trial.

International Registered Report Identifier (IRRID): RR2-10.2196/18688

(J Med Internet Res 2021;23(7):e27064) doi:10.2196/27064

KEYWORDS

telerehabilitation; teleassessment; arthrogryposis multiplex congenita; physical therapy; occupational therapy

Introduction

Background

Arthrogryposis multiplex congenita (AMC) is an umbrella term used to describe more than 400 conditions characterized by congenital joint contractures present in at least two body areas and affecting 1 in 4300-5100 live births [1,2]. Amyoplasia and distal arthrogryposis are the most common types of AMC, representing approximately 50%-65% of all AMC diagnoses [3,4]. Amyoplasia is characterized by decreases muscle mass and the typical positioning of the limbs (Figure 1) [5,6]. Distal arthrogryposis is characterized by contractures that mainly affect the distal limbs (ie, feet and hands; Figure 2) [6]. Individuals with AMC may have poor muscle mass in addition to limitations in range of motion (ROM), resulting in difficulties in transfers, mobility, and independent activities of daily living [6-8]. One aspect contributing to the limited ROM in individuals with AMC is the lack of movement of limbs, which may be overcome by physical exercise [6]. Some studies speculated that exercise in youth with AMC may maintain ROM, increase muscle strength, and decrease pain, but the direct effects of an exercise program remain to be tested in this population [3,6,9,10]. A positive association was also reported between knee and hip muscle strength and motor function (eg, rolling, sitting, standing, and climbing) in individuals with AMC, suggesting the importance of preserving sufficient muscle strength for daily activities [11]. Currently, most interventions in individuals with AMC, specifically rehabilitation, occur in early childhood, and the frequency decreases during school-age and adolescent years despite new challenges arising during these transition periods [12,13]. Rehabilitation for school-aged children with AMC focuses mostly on body functions (eg, mobility of joints and muscle endurance) and structure (eg, joint contracture of the elbow or knee), which does not always correspond to the individual's specific needs such as participating in activities and increasing independence for attending school and employment [14].

Figure 1. Young female with Amyoplasia with typical positioning of the limbs (ie, internal rotation at the shoulders, extension contractures at the elbow, and flexion contractures at the knees).



XSL•FO
RenderX

Figure 2. Adolescent with distal arthrogryposis, good range of motion at the shoulders and elbows (shoulder flexion and elbow flexion), and typical hand and foot deformities.



Rationale

Given the rarity, youths with AMC often require specialized care, but many of these individuals live far from subspecialized health care centers. Therefore, clinicians and researchers face challenges in developing adjunct therapies with frequent follow-ups, such as an exercise program. To overcome such challenges, novel intervention approaches and technologies are needed to increase access to subspecialized care for individuals with AMC living in remote areas. Telerehabilitation, an innovative way to deliver rehabilitation services remotely using telecommunication technologies, may offer a solution for delivering intervention programs with frequent follow-ups [15]. This new approach may contribute to reduced costs and decreased in-person visits to the hospital [16]. Telerehabilitation is not only useful for people with AMC living in remote areas but also useful for those living close by when they need consultations during a pandemic (such as the COVID-19 pandemic) in which social distancing is needed [17]. Despite its benefits, telerehabilitation has been understudied in children with physical disabilities [18]. Therefore, the primary aim of this pilot study is to evaluate the feasibility of using telerehabilitation to provide a home exercise program (HEP) for youth with AMC. As part of the feasibility objective, the interrater reliability of using a virtual goniometer for ROM measurements is explored. The secondary aim is to explore the effectiveness of this type of intervention.

Methods

Ethics and Recruitment

This prospective interventional single-cohort pilot study was conducted between January 2019 and March 2020 at the Shriners Hospitals for Children (SHC)–Canada. The study was approved by the department of medical research at the SHC (CAN1806) in July 2018, and ethics approval was received from the institutional review board of the McGill University Faculty of Medicine (#A08-B38-18B) in October 2018. Patients were included if they were aged between 8 and 21 years, understood written and spoken English or French, and had multiple congenital contractures as documented in their electronic medical records. Individuals were excluded if they had undergone a recent surgery (ie, 3 months prior for soft tissue and 6 months prior for bony surgery), lived outside Canada, or had cognitive deficits. Potential participants were identified by reviewing onsite medical records, and all eligible youths were approached during their clinic visit or by postal mail or phone if they did not have an upcoming appointment. Recruitment was also sought by posting an advertisement on a Canadian AMC support group on social media. Informed consent forms were filled by all parents and youths aged ≥14 years. Assent was provided for those aged 8-13 years.

Assessment

To establish the 12-week HEP, an initial assessment was conducted by an occupational therapist and a physical therapist using a videoconferencing platform. This assessment included

XSL•FO
RenderX

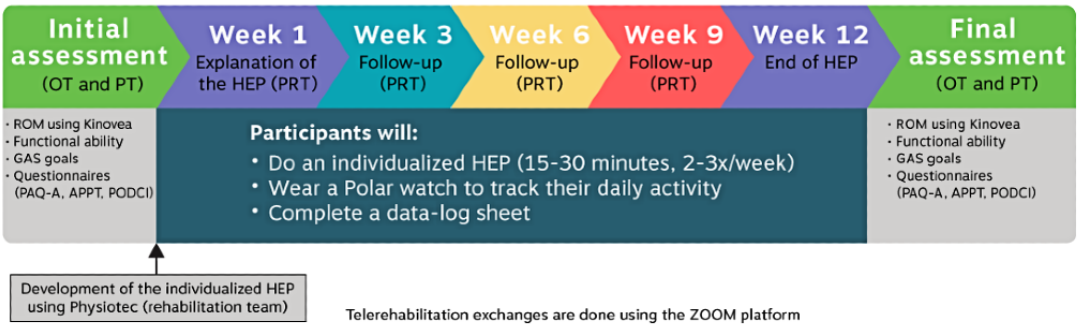
active ROM measurements of the upper and lower limb joints, an evaluation of overall function (mobility, transfers, and activities of daily living), and a pain assessment using the Adolescent and Pediatric Pain Tool (APPT). For the ROM assessment, the participants were asked to hold different positions, and screen captures were collected for subsequent measurements. The therapists guided the participants and the caregiver, when present, to ensure that the movements were performed properly and in the right plane of the camera (eg, the therapists asked for the camera to be tilted when needed). Based on the assessment and needs of the participants, the therapists and participants determined individualized goals using the Goal Attainment Scale (GAS). The scoring of the goals was defined according to the type of goal set, such as the percentage of perceived strength, a 0-10 satisfaction scale, or walking endurance measured in minutes. The information obtained through this initial assessment was used by the rehabilitation team to develop an individualized 12-week HEP. In addition, the participants were asked to complete web-based questionnaires, including the Physical Activity Questionnaire for Adolescents (PAQ-A) and the Pediatrics Outcomes Data Collection Instrument (PODCI), and to provide baseline disorder and sociodemographic information (ie, type of AMC, location of joint contractures, services care received, schooling, employment, volunteering, and leisure). All telerehabilitation meetings were conducted remotely using Zoom Pro (Zoom Video Communications Inc), a videoconferencing platform that allowed an encrypted connection and synchronous exchange between the participants and therapists. When needed, a parent was present to help position the camera and position or assist the participant or aid in the exercise material setup. In 1 case, the telerehabilitation meetings took place at the participant's school with the assistance of the school therapist. The methods used for the pilot study are described in detail in Gagnon et al [19].

At the end of the 12-week HEP, the participants underwent a final assessment, with the occupational therapist and physical therapist using the same outcome measures as those used during the initial assessment, and they completed the same web-based questionnaires with the addition of 19 closed-ended and 4 open-ended questions on their satisfaction with the HEP.

Intervention

All information gathered during the initial assessment was used to build an individualized HEP for each participant based on their goals and abilities. The therapists used Physiotec (Physiotec Québec Inc), a software program that easily creates an HEP and provides participants with access to detailed instructions and videos of the exercises. Examples of these HEPs can be found in the Multimedia Appendix 1. A week after the initial assessment, a physical rehabilitation therapist (PRT), a professional whose role is to develop treatment plans and provide appropriate intervention in collaboration with a physical therapist, explained the HEP to each participant during a remote session. When needed, the PRT demonstrated the exercises. During this session, the PRT ensured that the participant understood and performed the exercises safely. The participants were asked to perform their HEP 3 times a week for approximately 15 to 30 minutes at a time. Follow-ups were provided every 3 weeks (ie, at weeks 3, 6, and 9) by the PRT to ensure that the exercises were properly performed and to assess the level of difficulty of the HEP. Exercises were adjusted for progression or regression as needed (eg, number of repetitions and type of exercise). When necessary, for optimal loading, materials such as elastic resistance bands or theraputty (a silicone-based exercise material used for hand therapy) were sent to participants by postal mail. To record the compliance to the HEP, the participants were asked to wear a Polar A370 activity monitor (Polar Electro, Inc), which was sent to them by postal mail, to capture their exercise sessions and to complete a data log sheet. The activity monitor was worn on the wrist during the HEP session and was used to capture the number of exercise sessions. The participants were asked to charge their activity monitor every 3 to 4 days and, at the same time, to download their data to the Polar Flow platform through Bluetooth or direct connection. The downloaded data became available to the research team throughout the Polar Flow for Coach web platform. The research assistant (MG) manually scrutinized all recorded sessions on the Polar Flow for Coach web platform to remove those that were too short to be considered exercise sessions (≤15 min) or were recorded as another sport and made a note on the data log sheet (eg, hockey or skiing). Figure 3 shows a summary of the intervention.

Figure 3. Timeline of the telerehabilitation intervention. APPT: Adolescent and Pediatric Pain Tool; GAS: Goal Attainment Scale; HEP: home exercise program; OT: occupational therapist; PAQ-A: Physical Activity Questionnaire for Adolescents; PODCI: Pediatrics Outcomes Data Collection Instrument; PRT: physical rehabilitation therapist; PT: physical therapist; ROM: range of motion.



XSL•FO
RenderX

Statistical Analysis

Overview

Given that this was a pilot study with a small sample size and a heterogeneous population, nonparametric and descriptive statistics were used. *P* value corrections for multiple comparisons were not used because of the exploratory purpose of this study. Statistical analyses were performed using IBM SPSS Statistics version 24 for Windows (IBM Corp).

Feasibility

The feasibility was evaluated using different operationalization criteria. These operationalization criteria are listed in Textbox 1 and defined in detail in the published protocol for this study [19]. Compliance to the HEP and the telerehabilitation meetings were described using summary statistics and compared with pre-established feasibility criteria using a 1-sample Wilcoxon signed-rank test. For the remaining operationalization criteria, a comparison with pre-established criteria was performed using the same method. The technical issues (major technical issues being issues that resulted in the cancellation of the meeting and minor issues being something that could be resolved during the meeting) experienced during the program were also descriptively assessed using summary statistics. As part of evaluating feasibility, this pilot study also aimed to determine the most suitable outcome measures for this type of intervention.

Textbox 1. List of the operationalization criteria for assessing feasibility.

Source of recruitment

- At clinic, postal mail, phone, and social media

Recruitment rates

- ≥50% of the eligible and reachable youths

Withdrawal rates (before the intervention)

- ≤20% of the youths who consent

Withdrawal rates (during the course of the intervention)

- ≤30% of the youths who start the intervention

Completion rates

- ≥50% of the youths who consent

Compliance to the home exercise program

- ≥50% of compliance to the home exercise program trainings

Compliance to the telerehabilitation meetings

- ≤15% of the meetings missed

Missing data

- ≤10% for each outcome

Technical issues

- Echo voices, connection, and image quality

Interrater Reliability

The interrater reliability of using a virtual goniometer to measure active ROM has been shown to be feasible when compared with an in-person assessment with a goniometer on 10 healthy adults [20]. However, its reliability has not been assessed in children and adolescents with a musculoskeletal disorder such as AMC. The interrater reliability was established by having 2 raters (MG and GMM) use a virtual goniometer (Kinovea version 0.8.15) to measure the ROM of 4 participants selected at random. ROM measurements taken in the appropriate plane of movement were included in the analysis. The intraclass correlation coefficient (ICC) and associated 95% CI for each joint and overall were calculated based on a single-measurement, absolute-agreement, two-way random-effects model [21-23].

Effectiveness

Summary statistics, including the median and 95% CI for continuous variables and counts and proportions for categorical variables, were produced for all variables. As this was a pilot study on a rare disorder, statistical comparisons were produced only for exploratory purposes using a significance (α) level set a priori to 10% as has been proposed for research on rare disorders [24].

The effectiveness of the HEP was explored using the GAS. Raw GAS scores were converted to GAS *t* scores for each participant and the global HEP [25]. Within-participant pre- and postintervention scores of the PODCI, PAQ-A, APPT, and ROM were compared using the related-samples Wilcoxon signed-rank test. Individual changes were also compared to evaluate

XSL•FO
RenderX

clinically important differences as defined by Oeffinger et al [26] in the following domains of the PODCI (changes in points associated with a minimum clinically important difference): upper extremity (4.8 points), transfers and mobility (5.2 points), sports and physical function (10.3 points), pain and comfort (26.3 points), happiness (11.2 points), and global function (8.2 points). Changes in the APPT for the 10-point pain intensity scale were considered clinically meaningful when there was a change of more than two points [27,28].

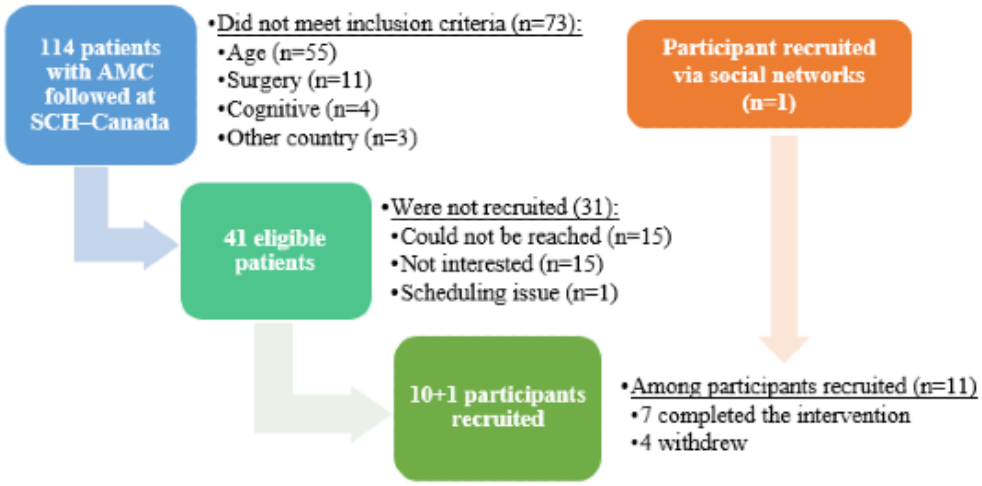
For the PODCI and PAQ-A questionnaires, a sensitivity analysis was conducted because a participant had an ankle injury, unrelated to the HEP execution, before the final assessment. An imputation technique was used to replace the final results of the participant who was directly affected by the injury by using the median of the group for the PAQ-A and in the following domains of the PODCI: transfer and mobility, sport and physical function, pain and comfort, and global function.

Results

Participant Information

Of the 114 patients with AMC followed at the SHC–Canada, 26 (22.8%) were eligible to participate and were able to be approached for participation. Of these, 10 individuals consented to the study, and an additional participant not actively followed at the SHC was recruited on social media. A total of 7 participants (5 boys) completed the intervention. The participants were recruited at the clinic (n=6) or through postal mail (n=2), phone (n=2), or social media (n=1). Of the 4 participants who withdrew, 2 consented at the clinic but never started the HEP because they could not be reached to schedule the initial assessment, and 2 withdrew after the start of the HEP because of personal reasons and lack of time. Figure 4 shows the participation flow diagram. The participants' median age was 16.9 years (range 11.3-20.8 years), and they represented 4 Canadian provinces with a median distance from the SHC–Canada of 227 km (range 7-3439 km). A total of 2 participants had Amyoplasia, 4 had distal arthrogryposis, and 1 had an unknown type of AMC. Joint involvement was distributed as follows: shoulders (n=5), elbows (n=6), wrists (n=6), hands (n=4), hips (n=3), knees (n=5), ankles (n=6), and spine (n=1), with a median of 6 joints (95% CI 3-8) involved. For the level of ambulation, 4 participants were classified as community ambulators, 2 were classified as household ambulators, and 1 was classified as nonambulatory.

Figure 4. Participant flowchart. AMC: arthrogryposis multiplex congenita.



Feasibility

The recruitment rate among eligible participants followed at the SHC who were reachable was 38% (10/26), which was lower than the target set at ≥50%. The withdrawal rate before the start of the intervention was 18% (2/11) and after the start of the intervention, it was 18% (2/11), which met the established criteria of ≤20% and ≤30%, respectively. The HEP completion rate was 64% (7/11), which corresponded to the criterion of ≥50%. The number of meetings that needed to be rescheduled in ≤24 hours was 11% (5/47), which reached the target set at ≤15%. The median HEP compliance was 2.04 times per week (95% CI 1.25-4.08), which when compared with the 3 weekly sessions set in the HEP, corresponded to 68.1% (95% CI 41.7%-136.1%) and was higher than the target set at ≥50% (P=.046). Different technical issues arose during the telerehabilitation meetings. No major technical issue necessitating canceling a meeting occurred, but minor technical problems occurred in 11% (5/47) of meetings. Echo voices were present with 1 participant, and this was resolved with the muting option: when the therapists or the participant spoke, the others muted themselves. The therapist lost the connection in one meeting for a few seconds and had to reconnect. On one occasion, a participant had difficulty with the sound output on their computer and had to join the meeting with their phone as an alternative. Regarding issues with the activity monitor, 5 participants had difficulties syncing the device, and the project coordinator resolved this problem by sharing the screen during the telerehabilitation meetings to review the required steps for setting it up.

All participants completed the pre- and postintervention questionnaires, except for 1 participant who did not return the data log sheet and never synced their activity monitor; therefore, they were excluded from the HEP compliance calculation despite reporting having performed their HEP regularly. For

XSL•FO
RenderX

the ROM measurements, the missing values for each joint can be found in Multimedia Appendix 2. All 7 participants reported being comfortable communicating through the telerehabilitation platform, and the clarity of the video and audio of the platform was acceptable. The participants felt supported during the 12-week program, reported reduced travel time compared with if they had to visit the SHC in person, and expressed interest in using telerehabilitation in the future. All were satisfied or very satisfied with the remote evaluation, delivery of the HEP, follow-ups, and the overall organization of the 12-week program. The participants suggested increasing the span of the program (n=1), to include an in-person follow-up visit halfway through the program (n=1), and to include outcome measures targeting the adolescent age group (n=1). A summary of the improvements reported by the participants in the open-ended questions can be found in Table 1.

Table 1. Summary of the improvements reported in the open-ended questions (n=7).

Improvements reported	Participants, n (%)
Daily activities (transfers, descending stairs, walking ability, and greater efficiency propelling a manual wheelchair)	3 (43)
Strength	3 (43)
Endurance	2 (29)
Mobility	1 (14)
Balance	1 (14)

Interrater Reliability

The interrater reliability varied among different joints, with an ICC of 0.985 (95% CI 0.980-0.989) for all joints combined. The lowest ICC was in forearm pronation with a median ICC of 0.252 (95% CI −0.477 to 0.75), and the highest was in shoulder extension with a median ICC of 0.998 (95% CI 0.983-1.00). The specific ICC for each joint is presented in Table 2.

Table 2. Intraclass correlation coefficient for each joint.

Joint	Number[a]	Median (95% CI)
Shoulder abduction	10	0.668 (0.072 to 0.908)
Shoulder flexion	4	0.915 (0.158 to 0.994)
Shoulder extension	5	0.998 (0.983 to 1)
Elbow flexion	12	0.988 (0.956 to 0.997)
Elbow extension	12	0.99 (0.965 to 0.997)
Forearm pronation	10	0.252 (−0.477 to 0.75)
Forearm supination	10	0.691 (0.069 to 0.917)
Wrist flexion	12	0.707 (0.277 to 0.904)
Wrist extension	12	0.858 (0.073 to 0.968)
Hip flexion	14	0.72 (0.315 to 0.901)
Hip extension	11	0.833 (0.484 to 0.952)
Hip internal rotation	11	0.975 (0.914 to 0.993)
Hip external rotation	12	0.971 (0.901 to 0.991)
Knee flexion	14	0.992 (0.958 to 0.998)
Knee extension	16	0.986 (0.961 to 0.995)
Ankle dorsiflexion	12	0.917 (0.749 to 0.975)
Ankle plantarflexion	12	0.878 (0.636 to 0.963)

[a]Number of data included for analysis, with a possible maximum of 16 (4 participants×2 sides×2 times).

Effectiveness

For the GAS, the objectives varied among the participants and were related to pain management (n=2); endurance in manual labor (n=1), writing (n=1), standing (n=1), or walking (n=2); sports (sledge hockey: n=1; karate: n=1); and daily activities (buttoning up shirt: n=1; transfer ability: n=2; controlling stair descent: n=1; self-propelling a wheelchair: n=1; self-feeding with utensils: n=1). The median number of goals established with the participants was 2 (range 1-3). A total of 15 goals were established, and 12 goals were achieved. The 3 goals that were not achieved included pain management (n=2) and walking endurance (n=1). The overall program *t* score was 74.85, whereas the median *t* score among participants was 56.21 (95%

XSL•FO
RenderX

CI 30-72.82); both scores are higher than the threshold of 50 points, meaning that overall, the participants achieved their goals.

The pre- and postintervention results from the different questionnaires are shown in Table 3. A statistically significant change was observed in the pain and comfort domain of the PODCI (*P*=.08). The number of participants with clinically important changes was as follows: upper extremity (improvement: n=2; decrease: n=1), transfers and mobility (improvement: n=3; decrease: n=2), sports and physical function (improvement: n=1; decrease: n=1), pain and comfort (improvement: n=1; decrease: n=0), happiness (improvement: n=2; decrease: n=1), and global function (improvement: n=2; decrease: n=0). For the PAQ-A, a statistically significant improvement of 14% was observed in the group (*P*=.046). Multimedia Appendices 3-4 show the pre- and postintervention results for each participant for the global domain of the PODCI and PAQ-A, respectively. For the APPT, 4 of 7 participants reported pain at baseline. One participant reported clinically meaningful improvement in the APPT at the end of the 12-week HEP; no changes were statistically significant.

For ROM, improvements in shoulder abduction (*P*=.08), shoulder flexion (*P*=.07), wrist extension (*P*=.05), and knee extension (*P*=.04) were found to be statistically significant. No significant changes were observed in the other joints. The results for ROM for each joint are shown in Table 4.

Table 3. Pre- and postintervention results from the different questionnaires.

Questionnaires	Preintervention, median (95% CI)	Postintervention, median (95% CI)	*P* value
PODCI[a]			
Upper extremity	87.50 (12.50 to 100)	95.83 (4.17 to 100)	.79
Transfer and mobility	93.94 (21.97 to 100)	91.66 (12.12 to 100)	.60
Sports and physical function	79.86 (11.36 to 86.11)	63.64 (13.64 to 93.18)	.61
Pain and comfort	71.11 (34.44 to 100)	85.00 (49.44 to 100)	*.08*[b]
Happiness	85.00 (50.00 to 100)	90.00 (35.00 to 100)	.50
Global function	87.47 (20.07 to 94.86)	82.08 (19.84 to 94.55)	.61
PAQ-A[c]	1.62 (1.00 to 2.82)	2.32 (1.00 to 3.45)	*.046*
APPT[d]			
Location (number)	1 (0 to 6)	1 (0 to 6)	>.99
Scale (cm)	1.55 (0 to 5.50)	1.10 (0 to 6.20)	.72
Sensory (%)	5.40 (0 to 10.8)	2.70 (0 to 10.8)	>.99
Affective (%)	0 (0 to 0)	0 (0 to 0)	>.99
Evaluative (%)	0 (0 to 25.00)	0 (0 to 37.5)	>.99

[a]PODCI: Pediatric Outcomes Data Collection Instrument.

[b]Italicized values indicate a significance level of *P*<.10.

[c]PAQ-A: Physical Activity Questionnaire for Adolescents.

[d]APPT: Adolescent Pediatric Pain Tool.

XSL•FO
RenderX

Table 4. Range of motion.

Joint	Data included for analysis[a]	Preintervention (in degrees), median (95% CI)	Postintervention (in degrees), median (95% CI)	*P* value
Shoulder abduction	12	143 (16 to 159)	151.5 (25 to 163)	*.08*[b]
Shoulder flexion	6	34.5 (0 to 152)	57.5 (0 to 153)	*.07*
Shoulder extension	3	66 (66 to 75)	60 (50 to 66)	.11
Elbow flexion	14	146.5 (86 to 164)	144.5 (92 to 158)	.67
Elbow extension	11	−9 (−82 to 0)[c]	−4 (−95 to 9)[c]	.35
Forearm pronation	9	82 (65 to 91)	95 (78 to 102)	.12
Forearm supination	9	71 (44 to 86)	66 (49 to 84)	.48
Wrist flexion	8	86.5 (29 to 97)	78 (11 to 94)	.11
Wrist extension	8	20.5 (−20 to 50)	30 (−7 to 57)	*.05*
Hip flexion	12	109.5 (95 to 127)	108.5 (94 to 130)	.92
Hip extension	8	1.5 (−50 to 30)	−1.5 (−45 to 16)[c]	.25
Hip internal rotation	3	1 (1 to 27)	12 (10 to 27)	.18
Hip external rotation	4	30 (4 to 55)	33 (11 to 48)	.72
Knee flexion	12	105.5 (50 to 130)	115 (93 to 138)	.61
Knee extension	14	−18 (−63 to −7)[c]	−14.5 (−62 to −4)[c]	*.04*
Ankle dorsiflexion	12	−21 (−28 to 5)[c]	−20.5 (−29 to −2)[c]	.72
Ankle plantarflexion	12	32 (23 to 40)	29.5 (21 to 41)	.53

[a]Number of data included for analysis, with a possible maximum of 14 (7 participants×2 sides).

[b]Italicized values indicate a significance level of $P<.10$.

[c]Negative values represent a lack of range of motion. For example, a negative knee extension signifies an inability to achieve full extension. Range of motion was measured to whole degrees; 0.5° resulted from median calculations of even numbers.

Discussion

Principal Findings

In this study on the use of telerehabilitation to deliver an HEP to youth with AMC, it was found that this approach is feasible and well accepted by participants. The GAS was a feasible measure in this pilot study because it provided the individualization of goals for the youths living with AMC who presented with varying levels of joint involvement. Overall, most of the GAS goals were achieved at the end of the 12-week HEP. In addition, some clinically meaningful improvements were observed in the PODCI and APPT scores, showing promising effects of an HEP. We also explored the reproducibility of using a virtual goniometer in this study and found good-to-excellent agreement overall.

Feasibility

Feasibility was demonstrated with the achievement of all operationalization criteria, except for the recruitment rates. The observed completion rate of 63.6% corresponded to the criteria set at ≥50%. One withdrawal was due to the need for more support from the caregiver to perform the HEP. Support from families or the entourage of the participant was important for most of the participants, specifically for the youngest participants and those with more severe joint involvement. The intervention in this study was semisupervised, with follow-ups provided every 3 weeks. The compliance of 68.1% (2.04/3 times per week) to the HEP measured in this study is closer to the compliance rate of 76% observed in supervised interventions [29] than to that of nonsupervised interventions (11%-37%) [30]. A total of 11% (5/47) of the telerehabilitation meetings were canceled on the same day, which is similar to the 12% (997/8306) observed for in-person appointments in the rehabilitation department at the SHC–Canada (unpublished data, 2019). Extreme weather in Canada often results in cancellation of appointments, but none were canceled for this reason in this study owing to its web-based care delivery. A few technical issues occurred during this study, but all were resolved promptly, and none affected the safety of the participants. Therefore, technical issues should not be a reason to preclude the use of telerehabilitation in future studies or in clinical practice.

Although the reproducibility of using a virtual goniometer varied across joints, this method of measuring ROM was nevertheless useful for setting individualized goals, developing the HEP, and performing the initial and final assessments. Some challenges with web-based ROM measurements included different movement planes for measurement, varying positions in which ROM measurements were taken according to the participants' contractures (eg, elbow flexion in sitting or standing position), space restrictions in the home, or poor or high luminosity. In future studies using web-based ROM measurements, we recommend the use of guidelines for proper positioning,

XSL•FO
RenderX

therapist training, contrasting clothing for the participants, proper luminosity, and laptop or tablet use to allow adjustment of the camera as needed.

Effectiveness

The results regarding the intervention's effectiveness are promising. The GAS score, which was the main outcome measure to explore effectiveness, showed significant improvement at the end of the 12-week HEP. Among the 12 goals achieved measured using the GAS, 9 were achieved beyond expectations: endurance in manual labor, standing, and walking; sports (karate); and daily activities (buttoning up shirt, transfer ability, controlling stair descent, self-propelling a wheelchair, and using a fork). These large improvements might be due to the scaling of the goals using the GAS table. As most of the participants did not perform exercises before participating in this project, they may have demonstrated quicker gains. However, 2 participants with chronic pain did not achieve their pain goal, perhaps because of the multifactorial etiology of pain requiring management by an interdisciplinary team [31]. Another participant did not meet their initial goal for walking endurance as per the GAS score because of overestimation of baseline levels as reported by the participant; nonetheless, improvements during the final assessment (quicker walking pace) were noted. As the GAS allows for individualized and varied goals, it proved to be a versatile and sensitive measurement tool, given the high level of variability of joint involvement and physical function. There are benefits to using the GAS, but this tool requires some level of training and experience to develop goals that are SMART (specific, measurable, achievable, realistic and timed) [32]. After determining individualized goals in collaboration with participants, therapists need to scale the goals according to an expected level of change and predict potential improvement within the timeline, which may not always be intuitive.

Although some clinically significant changes were measured with the PODCI, floor and ceiling effects were observed in a few participants. For example, in a participant with severe joint involvement, a floor effect was noted in all but the happiness domains. Large improvements in the transfer ability from wheelchair to exercise table in this participant were noted during the telerehabilitation sessions, yet the PODCI did not detect this change because the focus of the PODCI is mainly on ambulatory activities such as walking and running. The opposite effect occurred in participants who achieved maximum scores at baseline and, therefore, no positive change was detected. For these reasons, we conclude that the PODCI may not be an appropriate measure for our study, which used an exercise intervention for a heterogeneous group of youths. The baseline results from the PAQ-A were very low, yet they increased at the end of the 12-week HEP, with levels approaching those of typically developing adolescents [33]. The level of physical activity among participants with AMC remained lower than that among typically developing adolescents. This was similar to a study of adolescents with cerebral palsy [34]. With regard to ROM, we did not expect any significant changes because no participant selected ROM as their goal. Some statistically significant improvements were found, and they may have been explained by an improvement in muscle strength; however, these changes may not translate into clinically important changes.

Limitations

Most of the outcomes were standardized, although the HEP was based on individualized goals. The assessment provided did not always correspond to the needs of the participants because they did not provide objective information related to the individualized goals (eg, strength or endurance). For example, joint-specific ROM was measured for all participants, but it was only pertinent for 1 participant who had a dressing goal. For the interrater reliability of ROM measurement with a virtual goniometer, it would be best to compare those values with in-person measurements. However, in the context of this study, it was not possible to do so because the study design entailed remote visits only. Therefore, the ROM results should be interpreted with caution. The selected questionnaires for this study were validated for children aged up to 18 years, which corresponded to most of the participants included in this study (5/7, 71%). However, at the SHC–Canada, patients are followed until 21 years of age, and an older participant mentioned that the content of some of the questionnaires was not adapted to their reality. Finding standardized measures devised and validated for both pediatric and young adult populations is a challenge because measures are typically developed for either pediatric or adult populations. As this was a pilot study with the main objective being the assessment of its feasibility, no control group was included. A larger multicenter randomized controlled trial would be needed to ascertain the study power needed to draw conclusions on the effectiveness of an HEP. Initial and final assessments performed in person with follow-ups conducted remotely could also be a good compromise for future studies.

Conclusions

The results of this pilot study suggest that it is feasible to use telerehabilitation to deliver an HEP for youth with AMC having different functional levels, having various goals, and living across different geographical regions. This approach also holds promise regarding the effectiveness of an HEP. Lessons can be learned from the challenges and positive aspects denoted in this study, which could lead to enhanced future protocols.

Acknowledgments

This research was funded by the Canadian Musculoskeletal Rehab Research Network (CIHR FRN CFI-148081), Can $10,000 (US $8289). MG was awarded a master's scholarship from the Grand Défi Pierre Lavoie Foundation, Can $7500 (US $6217). ND received salary support from Fonds de recherche du Québec-Santé. The authors owe special thanks to all the youths and their families for their participation in this project. They acknowledge the contribution of Dr Emmanouil Rampakakis for his advice on statistical analysis and review of the manuscript. They also thank Guylaine Bédard from the SHC–Canada for the illustration.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
An exercise program is provided as an example for each participant.
[PDF File (Adobe PDF File), 4947 KB - jmir_v23i7e27064_app1.pdf]

Multimedia Appendix 2
This table shows the missing values and their associated reasons for each range of motion measurement.
[DOCX File , 18 KB - jmir_v23i7e27064_app2.docx]

Multimedia Appendix 3
A bar graph that shows the pre- and postintervention global function scores of the Pediatric Outcomes Data Collection Instrument for each participant and the clinically significant difference. *Medium clinically significant difference, ** Large clinically significant difference.
[PNG File , 15 KB - jmir_v23i7e27064_app3.png]

Multimedia Appendix 4
A bar graph that shows the pre- and postintervention Physical Activity Questionnaire for Adolescents results for each participant. *Difference between 10% and 19%; **difference between 20% and 29%; ***difference ≥30%.
[PNG File , 13 KB - jmir_v23i7e27064_app4.png]

References

1. Lowry RB, Sibbald B, Bedard T, Hall JG. Prevalence of multiple congenital contractures including arthrogryposis multiplex congenita in Alberta, Canada, and a strategy for classification and coding. Birth Defects Res A Clin Mol Teratol 2010 Dec 15;88(12):1057-1061. [doi: 10.1002/bdra.20738] [Medline: 21157886]
2. Dahan-Oliel N, Cachecho S, Barnes D, Bedard T, Davison AM, Dieterich K, et al. International multidisciplinary collaboration toward an annotated definition of arthrogryposis multiplex congenita. Am J Med Genet C Semin Med Genet 2019 Sep;181(3):288-299 [FREE Full text] [doi: 10.1002/ajmg.c.31721] [Medline: 31282072]
3. Hall J. Arthrogryposis (multiple congenital contractures): diagnostic approach to etiology, classification, genetics, and general principles. Eur J Med Genet 2014 Aug;57(8):464-472. [doi: 10.1016/j.ejmg.2014.03.008] [Medline: 24704792]
4. Hall J, Kimber E, van Bosse HJ. Genetics and classifications. J Pediatr Orthop 2017;37 Suppl 1:4-8. [doi: 10.1097/BPO.0000000000000997] [Medline: 28594686]
5. Hall J, Reed S, Driscoll E. Part I. Amyoplasia: a common, sporadic condition with congenital contractures. Am J Med Genet 1983 Aug;15(4):571-590. [doi: 10.1002/ajmg.1320150407] [Medline: 6614047]
6. Staheli L, Hall J, Jaffe KM, Paholke DO. Arthrogryposis: A Text Atlas. United Kingdom: Cambridge University Press; 1999:A.
7. Dubousset J, Guillaumat M. Long-term outcome for patients with arthrogryposis multiplex congenita. J Child Orthop 2015 Dec;9(6):449-458 [FREE Full text] [doi: 10.1007/s11832-015-0692-6] [Medline: 26499455]
8. Dillon E, Bjornson K, Jaffe K, Hall J, Song K. Ambulatory activity in youth with arthrogryposis: a cohort study. J Pediatr Orthop 2009 Mar;29(2):214-217. [doi: 10.1097/BPO.0b013e3181990214] [Medline: 19352250]
9. Rodrigues E, Gomes A, Tanhoffer A, Leite N. Effects of exercise on pain of musculoskeletal disorders: a systematic review. Acta Ortop Bras 2014;22(6):334-338 [FREE Full text] [doi: 10.1590/1413-78522014220601004] [Medline: 25538482]
10. Skalsky AJ, McDonald CM. Prevention and management of limb contractures in neuromuscular diseases. Phys Med Rehabil Clin N Am 2012 Aug;23(3):675-687 [FREE Full text] [doi: 10.1016/j.pmr.2012.06.009] [Medline: 22938881]
11. Kroksmark A, Kimber E, Jerre R, Beckung E, Tulinius M. Muscle involvement and motor function in amyoplasia. Am J Med Genet A 2006 Aug 15;140(16):1757-1767. [doi: 10.1002/ajmg.a.31387] [Medline: 16835916]
12. Gagnon M, Caporuscio K, Veilleux LN, Hamdy R, Dahan-Oliel N. Muscle and joint function in children living with arthrogryposis multiplex congenita: a scoping review. Am J Med Genet C Semin Med Genet 2019 Sep;181(3):410-426. [doi: 10.1002/ajmg.c.31726] [Medline: 31350946]
13. Elfassy C, Darsaklis V, Snider L, Gagnon C, Hamdy R, Dahan-Oliel N. Rehabilitation needs of youth with arthrogryposis multiplex congenita: perspectives from key stakeholders. Disabil Rehabil 2020 Aug;42(16):2318-2324. [doi: 10.1080/09638288.2018.1559364] [Medline: 30741031]
14. Wagner LV, Cherry JS, Sawatzky BJ, Fąfara A, Elfassy C, Eriksson M, et al. Rehabilitation across the lifespan for individuals with arthrogryposis. Am J Med Genet C Semin Med Genet 2019 Sep 08;181(3):385-392. [doi: 10.1002/ajmg.c.31729] [Medline: 31397084]

XSL•FO
RenderX

15. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: a noninferiority randomized controlled trial. J Bone Joint Surg Am 2015 Jul 15;97(14):1129-1141. [doi: 10.2106/JBJS.N.01066] [Medline: 26178888]
16. Tousignant M, Moffet H, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. Cost analysis of in-home telerehabilitation for post-knee arthroplasty. J Med Internet Res 2015 Mar 31;17(3):e83 [FREE Full text] [doi: 10.2196/jmir.3844] [Medline: 25840501]
17. Keshet D, Bernstein M, Dahan-Oliel N, Ouellet J, Pauyo T, Rabau O, et al. Management of common elective paediatric orthopaedic conditions during the COVID-19 pandemic: the Montreal experience. J Child Orthop 2020 Jul 01;14(3):161-166 [FREE Full text] [doi: 10.1302/1863-2548.14.200107] [Medline: 32582382]
18. Camden C, Pratte G, Fallon F, Couture M, Berbari J, Tousignant M. Diversity of practices in telerehabilitation for children with disabilities and effective intervention characteristics: results from a systematic review. Disabil Rehabil 2020 Dec;42(24):3424-3436. [doi: 10.1080/09638288.2019.1595750] [Medline: 30978110]
19. Gagnon M, Collins J, Elfassy C, Merlo GM, Marsh J, Sawatzky B, et al. A telerehabilitation intervention for youths with arthrogryposis multiplex congenita: protocol for a pilot study. JMIR Res Protoc 2020 Jul 26;9(6):e18688 [FREE Full text] [doi: 10.2196/18688] [Medline: 32589157]
20. Durfee WK, Savard L, Weinstein S. Technical feasibility of teleassessments for rehabilitation. IEEE Trans Neural Syst Rehabil Eng 2007 Mar;15(1):23-29. [doi: 10.1109/tnsre.2007.891400]
21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012]
22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1(1):30-46. [doi: 10.1037/1082-989x.1.1.30]
23. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420]
24. Bogaerts J, Sydes MR, Keat N, McConnell A, Benson A, Ho A, et al. Clinical trial designs for rare diseases: studies developed and discussed by the International Rare Cancers Initiative. Eur J Cancer 2015 Mar;51(3):271-281 [FREE Full text] [doi: 10.1016/j.ejca.2014.10.027] [Medline: 25542058]
25. Kiresuk T, Smith A, Cardillo J. Goal Attainment Scaling: Applications, Theory, and Measurement. London: Psychology Press; 2014:A.
26. Oeffinger D, Bagley A, Rogers S, Gorton G, Kryscio R, Abel M, et al. Outcome tools used for ambulatory children with cerebral palsy: responsiveness and minimum clinically important differences. Dev Med Child Neurol 2008 Dec;50(12):918-925 [FREE Full text] [doi: 10.1111/j.1469-8749.2008.03150.x] [Medline: 19046185]
27. Farrar J, Young JJ, LaMoreaux L, Werth J, Poole M. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Dec;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9] [Medline: 11690728]
28. Tsze DS, Hirschfeld G, von Baeyer CL, Suarez LE, Dayan PS. Changes in pain score associated with clinically meaningful outcomes in children with acute pain. Acad Emerg Med 2019 Sep 19;26(9):1002-1013 [FREE Full text] [doi: 10.1111/acem.13683] [Medline: 30636350]
29. Wong YK, Hui E, Woo J. A community-based exercise programme for older persons with knee pain using telemedicine. J Telemed Telecare 2005 Sep 01;11(6):310-315. [doi: 10.1258/1357633054893346] [Medline: 16168168]
30. Hartman A, te Winkel M, van Beek R, de Muinck S, Kemper H, Hop W, et al. A randomized trial investigating an exercise program to prevent reduction of bone mineral density and impairment of motor performance during treatment for childhood acute lymphoblastic leukemia. Pediatr Blood Cancer 2009 Jul 15;53(1):64-71. [doi: 10.1002/pbc.21942] [Medline: 19283791]
31. Manchikanti L, Singh V, Kaye A, Hirsch J. Lessons for better pain management in the future: learning from the past. Pain Ther 2020 Dec;9(2):373-391 [FREE Full text] [doi: 10.1007/s40122-020-00170-8] [Medline: 32410070]
32. Bovend'Eerdt T, Botell R, Wade D. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. Clin Rehabil 2009 May;23(4):352-361. [doi: 10.1177/0269215508101741] [Medline: 19237435]
33. Fernández I, Canet O, Giné-Garriga M. Assessment of physical activity levels, fitness and perceived barriers to physical activity practice in adolescents: cross-sectional study. Eur J Pediatr 2017 Jan;176(1):57-65. [doi: 10.1007/s00431-016-2809-4] [Medline: 27858223]
34. Maher C, Williams M, Olds T, Lane A. Physical and sedentary activity in adolescents with cerebral palsy. Dev Med Child Neurol 2007 Jul;49(6):450-457 [FREE Full text] [doi: 10.1111/j.1469-8749.2007.00450.x] [Medline: 17518932]

Abbreviations

AMC: arthrogryposis multiplex congenita
APPT: Adolescent and Pediatric Pain Tool
GAS: Goal Attainment Scale
HEP: home exercise program
ICC: intraclass correlation coefficient

XSL•FO
RenderX

PAQ-A: Physical Activity Questionnaire for Adolescents
PODCI: Pediatrics Outcomes Data Collection Instrument
PRT: physical rehabilitation therapist
ROM: range of motion
SHC: Shriners Hospitals for Children
SMART: specific, measurable, achievable, realistic, and timed

Edited by R Kukafka; submitted 10.01.21; peer-reviewed by H vanBosse, A Hidki, K Ndlovu; comments to author 16.03.21; revised version received 13.04.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Gagnon M, Marino Merlo G, Yap R, Collins J, Elfassy C, Sawatzky B, Marsh J, Hamdy R, Veilleux LN, Dahan-Oliel N
Using Telerehabilitation to Deliver a Home Exercise Program to Youth With Arthrogryposis: Single Cohort Pilot Study
J Med Internet Res 2021;23(7):e27064
URL: https://www.jmir.org/2021/7/e27064
doi:10.2196/27064
PMID:34255680

©Marianne Gagnon, Gabriela Marino Merlo, Rita Yap, Jessica Collins, Caroline Elfassy, Bonita Sawatzky, Jacquelyn Marsh, Reggie Hamdy, Louis-Nicolas Veilleux, Noémi Dahan-Oliel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of Telemedicine Services Based on a Regional Telemedicine Platform in China From 2014 to 2020: Longitudinal Trend Analysis

Fangfang Cui[1*], MM; Xianying He[1*], MM; Yunkai Zhai[2], PhD; Minzhao Lyu[3], BE (Hons); Jinming Shi[1], MSc; Dongxu Sun[1], MSc; Shuai Jiang[1], PhD; Chenchen Li[1], MM; Jie Zhao[1], PhD

[1]National Engineering Laboratory for Internet Medical Systems and Applications, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[2]School of Management Engineering, Zhengzhou University, Zhengzhou, China

[3]School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Jie Zhao, PhD
National Engineering Laboratory for Internet Medical Systems and Applications
The First Affiliated Hospital of Zhengzhou University
1 Jianshe Road, Erqi District
Zhengzhou
China
Phone: 86 371 67966215
Fax: 86 371 67966215
Email: zhaojie@zzu.edu.cn

Abstract

Background: Telemedicine that combines information technology and health care augments the operational model of traditional medical services and brings new opportunities to the medical field. China promotes telemedicine with great efforts, and its practices in the deployment of telemedicine platforms and delivery of services have become important references for the research and development in this field.

Objective: Our work described in this paper focuses on a regional telemedicine platform that was built in 2014. We analyzed the system design scheme and remote consultations that were conducted via the system to understand the deployment and service delivery processes of a representative telemedicine platform in China.

Methods: We collected information on remote consultations conducted from 2015 to 2020 via the regional telemedicine platform that employs a centralized architectural system model. We used graphs and statistical methods to describe the changing trends of service volume of remote consultation, geographical and demographic distribution of patients, and waiting time and duration of consultations. The factors that affect consultation duration and patient referral were analyzed by multivariable linear regression models and binary logistic regression models, respectively. The attitudes toward telemedicine of 225 medical practitioners and 225 patients were collected using the snowball sampling method.

Results: The regional telemedicine platform covers all levels of medical institutions and hospitals in all 18 cities of Henan Province as well as some interprovince hospitals. From 2015 to 2020, 103,957 remote medical consultations were conducted via the platform with an annual increasing rate of 0.64%. A total of 86.64% (90,069/103,957) of medical institutions (as clients) that applied for remote consultations were tier 1 or 2 and from less-developed regions; 65.65% (68,243/103,945) of patients who applied for remote consultations were aged over 50 years. The numbers of consultations were high for departments focusing in the treatment of chronic diseases such as neurology, respiratory medicine, and oncology. The invited experts were mainly experienced doctors with senior professional titles. Year of consultation, tier of hospital, consultation department, and necessity of patient referral were the main factors affecting the duration of consultations. In surveys, we found that 60.4% (136/225) of medical practitioners and 53.8% (121/225) of patients had high satisfaction and believed that telemedicine is of vital importance for the treatment of illness.

Conclusions: The development of telemedicine in China shows a growing trend and provides great benefits especially to medical institutions located in less developed regions and senior citizens who have less mobility. Cases of remote consultations are mainly

XSL•FO
RenderX

for chronic diseases. At present, the importance and necessity of telemedicine are well recognized by both patients and medical practitioners. However, the waiting time needs to be further reduced to improve the efficiency of remote medical services.

(J Med Internet Res 2021;23(7):e28009) doi:10.2196/28009

KEYWORDS

telemedicine; regional telemedicine service platform; remote consultation; efficiency; satisfaction degree; telehealth; mobile health; mHealth; remote; China

Introduction

Medical services are of vital importance for the community. In China, the unbalanced distribution of health care resources has become an urgent problem [1-4]. Medical resources between hospitals are not effectively shared. High-quality services are mainly offered in capital cities and developed regions, while patients from less-developed regions (eg, rural areas) may not receive treatments responsive to their needs. In addition, the limited resources provided by less capable hospitals may not fully satisfy the needs of patients with intractable and rare diseases. Some patients may visit large hospitals that provide better services. As a consequence, patients are under more financial pressure [5], and resources from small hospitals may not be properly used. Generally speaking, low-tier hospitals usually do not have the capability to provide high-quality medical services, while large hospitals in developed cities are overloaded with patients.

To address those pain points, in recent years there has been collaboration of Chinese medical services using the internet and information technology [6-8]. Such emerging service models provide solid supports for collaborations between medical institutions. For example, regional telemedicine platforms such as the Golden Health telemedicine network and the People's Liberation Army telemedicine network [9] have been deployed in many provinces and serve medical institutions at all levels [10,11], which significantly improves the service quality of primary hospitals and eases the pressure of overloading service volume on large hospitals. After the outbreaks of COVID-19, telemedicine platforms have been adopted by a large number of hospitals to conduct remote consultations (ie, teleconsultations), treatments, and ward rounds that reduce contacts between medical practitioners and patients [12-14].

On a global scale, telemedicine connects medical service providers in sparsely populated areas. Large or specialized hospitals offer remote consultations, diagnoses, and treatments to medical institutions and patients in less-developed regions [15]. In the early stage, telemedicine services were often performed using visual telephone, email, or Integrated Services Digital Network [16]. Later, conferencing software such as voice over Internet Protocol or FaceTime was used. In recent years, private medical networks and computing platform technologies gradually became the major enabling channels for telemedicine. In Albania, an open and shared telemedicine collaboration platform that has been widely used in Europe uses open-source technology and establishes connections via the internet [17]. Crespo et al [18] designed a telemedicine platform that provides remote health care services for elderly patients with chronic obstructive pulmonary disease. Beer et al [19] performed effective remote diagnosis and treatment of skin diseases through their telemedicine platform during the outbreak of COVID-19. In 2007, a regional telemedicine center that enables sharing of digital health care records and remote consultations between medical institutions was established in Gansu Province [10]. The West China Hospital has established a medical information platform that serves western China through digital networks and video equipment [11].

Telemedicine platforms achieve effective real-time collaborations between medical institutions from different regions or countries and cover areas including pediatric problems [20,21], skin diseases [22], neurosurgery [23], and diabetes [24]. Therefore, the service volume of telemedicine has increased dramatically. From 2005 to 2012, a state tertiary pediatric hospital in Western Australia provided remote treatments to 1312 children with burns [25]. The University of Rochester has established a telemedicine platform for the treatment of mental illness that conducts about 2000 telepsychiatric consultations per year [26]. In the United States, 15 million people were supported by telemedicine in 2015, an increase of 50% compared with the number in 2013 [27]. In China, the average number of remote consultations from hospitals that provide telemedicine services reached 714 cases per hospital in 2018 [28].

Existing research works mainly focus on the development of telemedicine platforms or analysis of medical techniques (eg, artificial intelligence–assisted diagnosis) toward a certain type of disease. However, to the best of our knowledge, few works analyze the process of telemedicine services to study how to improve the efficiency of services and satisfaction of involved parties (eg, medical practitioners and patients), critical for the optimization of telemedicine applications. In this paper, we described our research and focus on valuable insights obtained from the largest regional telemedicine platform in China, which connects more than 1000 registered medical institutions across the country. First, we introduced the design principle and choices of the platform. Second, we comprehensively analyzed the volume, process, and effectiveness of remote medical services supported by the platform. Our results provide important references for the optimization of service processes, increase in efficiency, and enhancement of service value.

Methods

Design of the Telemedicine Platform

The regional telemedicine platform is designed to be compatible with multiple network access methods. It connects 1037 medical institutions at the provincial, municipal, county, and township levels and provides telemedicine services to medical institutions

XSL•FO
RenderX

not only in Henan Province but also in other provinces of China (Multimedia Appendix 1). The platform provides services between hospitals (ie, business-to-business); it does not provide connections between medical institutions and patients at home (ie, business-to-consumer). Its capabilities include supporting communications via video terminal devices of diverse types, facilitating data sharing among medical institutions at all levels, and enabling remote medical services such as consultation and diagnosis. The platform is implemented using a centralized system architecture with modular design, where the central module (ie, the core of our system and information exchange) mainly contains the medical information management block and the Session Initiation Protocol service block. For information exchange, the platform uses web services and visualization of databases to achieve real-time sharing of medical data between the telemedicine center and connected hospitals, which makes it convenient for medical practitioners to view health records of patients during a consultation. For resource management, the platform adopts an Internet Protocol–based multimedia system that centrally manages resources of telemedicine networks, patients, medical practitioners, and services so that partner institutes are able to access from networks with heterogeneous types. The consultation and diagnosis systems that operate on the top of the platform are implemented using browser/server architecture, which supports common operating systems such as Windows, MacOS, and Linux. Using web terminals, platform managers can perform regular maintenance of key information including registered hospitals and medical experts. Furthermore, the invited experts can view electronic medical records and communicate with the host doctors via Session Initiation Protocol video conferencing.

Data Collection

We collected data from all medical cases that were consulted through the telemedicine platform from January 2015 to December 2020. Our dataset contains the geolocation of hospitals applied for remote consultations, gender and age information of patients, waiting times, consultation durations, withdrawals of consultation, and advice on patient referrals after teleconsultations. To guarantee the validity and effectiveness of our collected data, qualitative information on patients, medical practitioners, and consultation results is approved by all participants before being formally recorded. Quantitative data such as waiting time and consultation duration are automatically measured by the platform to avoid human error. We note that the waiting time is the period between an application being submitted by a partner hospital and the start time of the consultation. The consultation duration is the effective communication period between all involved parties of a consultation. We performed data cleansing and preprocessing to obtain a structured dataset that facilitates our analysis. In addition, to understand the effectiveness of remote consultations, we collected information on the attitudes toward each consultation from the participating patients and medical practitioners. To be more specific, we selected one representative medical institution in each city that is covered by our platform to conduct our survey. Both patients and medical practitioners were invited through the snowball sampling method. As the final step of their teleconsultation process, all invited participants answered our survey. From patients, we collected their opinions on level of satisfaction toward the consultation process and results, benefits of remote consultation in reducing financial costs, and increased convenience. We collected attitudes from medical practitioners regarding each consultation process and its effectiveness.

Statistical Analysis

During this study we collected data from 103,957 remote consultation cases. Numerical indicators such as quantity, average value, and composition ratio were used in our descriptive analysis. The quantitative data were described by the mean values, while the qualitative data were described by the counts and percentages. We calculated the average annual growth rate of teleconsultation encounters. With the support of Excel (Microsoft Corp) software, we drew line charts to describe the changing trends of consultation cases over time. Using SPSS (version 23.0, IBM Corp) software, we employed multivariable linear regression to analyze the impact of the information of applicant hospitals, patients, and consultations on consultation durations. Binary logistic regression analysis was used to understand the impact of the rank of applicant hospital, geolocation, patient status, consultation departments, and other factors on the referral recommendations. In our statistical tests, the significance levels were set to α=.05. Incorrect and incomplete data were treated as missing values.

Results

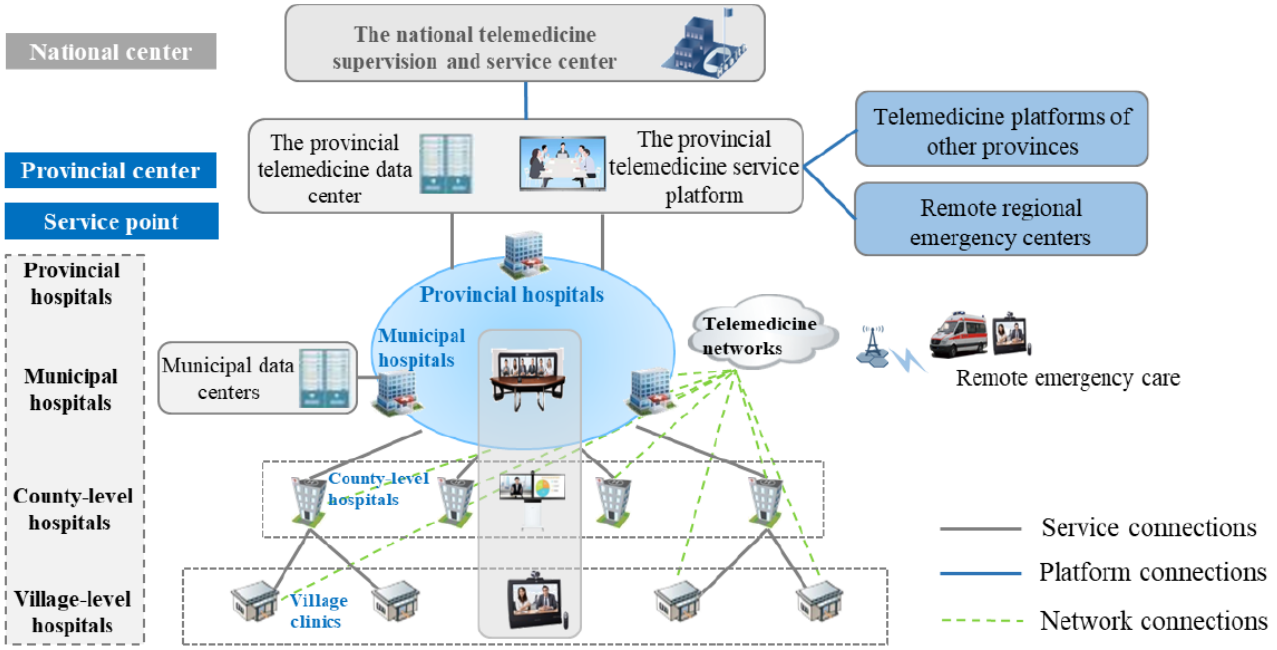
Regional Telemedicine Platform

The regional telemedicine platform connects medical institutions at all levels including provincial hospitals, municipal hospitals, county-level medical institutions, township health centers, community service centers, and village-level clinics. Figure 1 shows the overall connectivity scheme of the platform. Communications between each institution are established via virtual private medical networks that enable service collaboration and health data sharing across hospitals. Apart from the provincial center, municipal telemedicine centers are deployed in 18 cities of the Henan Province. Telemedicine centers of both levels (ie, provincial and municipal) have their own dedicated medical data centers, which together form a dual-active data center. Medical data that are used by and generated from each consultation are stored in the provincial data center and copied to the corresponding municipal data center. Medical institutions at various levels select different types of video conferencing equipment according to their needs and available resources. For example, provincial and municipal hospitals mainly use large-scale multiscreen consultation terminals with the highest audio and video quality, county-level institutions usually choose dual-screen separate consultation terminal devices, and township health centers and village-level clinics are likely to use single-screen computers with embedded cameras and microphones as their consultation terminals. The regional telemedicine platform is connected with the regional emergency center and telemedicine platform in other provinces via virtual private medical networks. In addition, mobile ambulances are connected to the platform through 5G networks.

XSL•FO
RenderX

A registered hospital can initiate a consultation by logging onto the platform to provide information on patients, invite participants, and submit an application. The invited hospital schedules the consultation and informs the corresponding medical practitioners. All participants join in the teleconsultation using terminal devices. Results of each consultation are automatically recorded and maintained by the system for future reference.

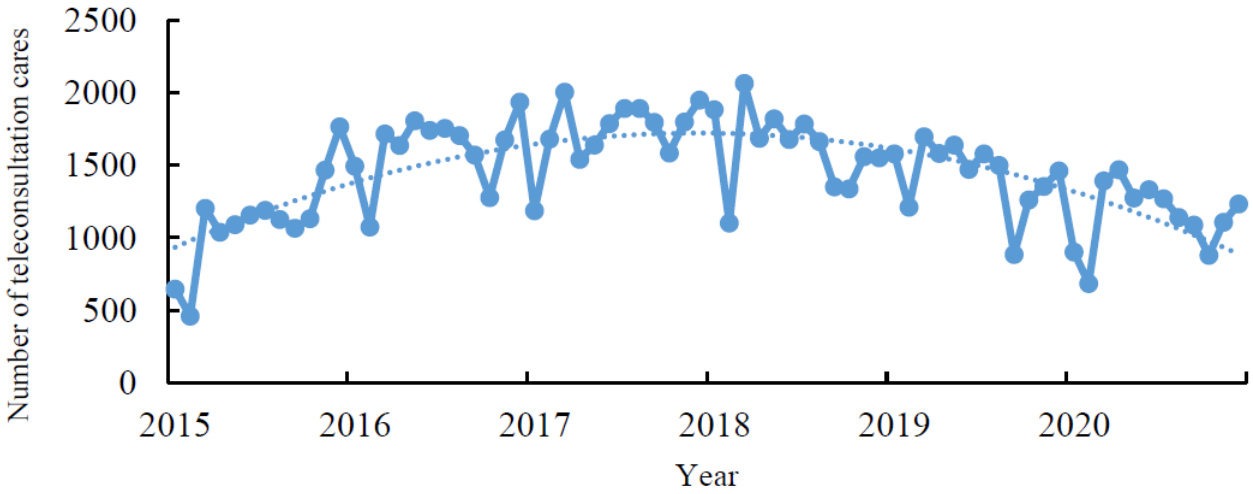
Figure 1. Overall connectivity of the regional telemedicine platform.



Teleconsultation Service Volume and Characteristics

From 2015 to 2020, 103,957 remote medical consultations were conducted through the telemedicine platform. There were 12.83% (13,337/103,957), 18.65% (19,391/103,957), 19.96% (20,753/103,957), 18.74% (19,480/103,957), 16.57% (17,225/103,957), and 13.25% (13,771/103,957) teleconsultation encounters in the years 2015, 2016, 2017, 2018, 2019, and 2020, respectively (Multimedia Appendix 2). There was an overall increasing trend with an annual rate of 0.64%. We observed fluctuations in the patterns in different months. The number of teleconsultation encounters was high in November and December and relatively low in January and February (Figure 2). That is, the service volume was large in the winter but smaller during Lunar New Year (ie, January and February).

Figure 2. Changing tendency of service volume from 2015 to 2020.

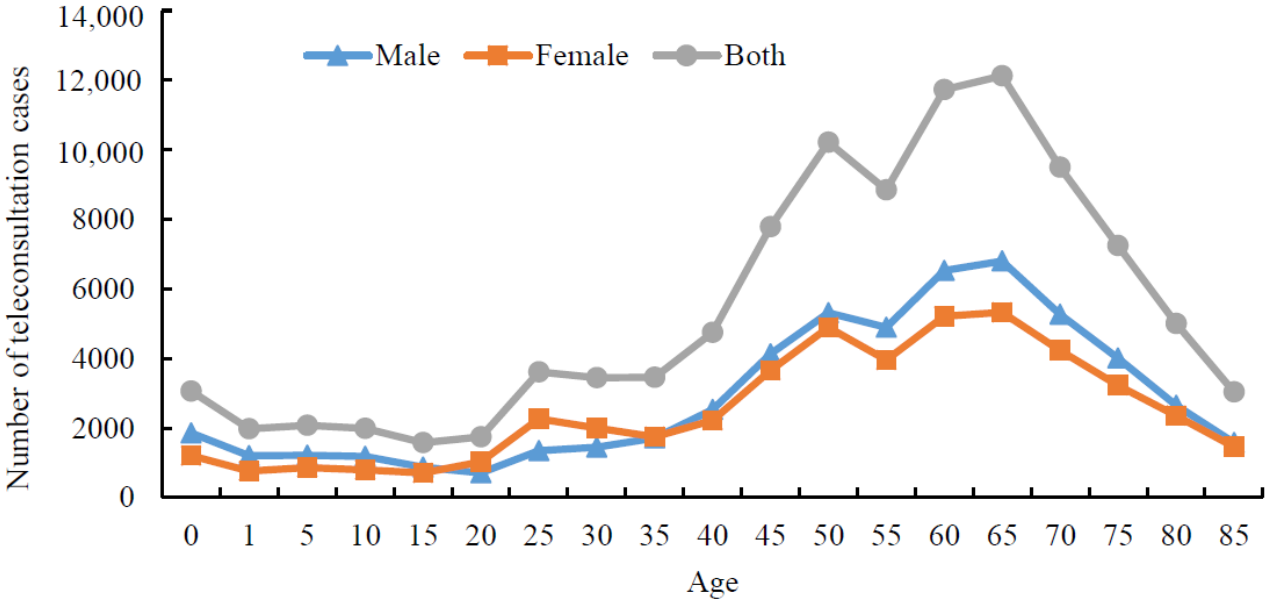


Considering the characteristics of patients who participated in teleconsultation encounters, 53.55% (55,664/103,945) were male. The majority (68,243/103,945, 65.65%) were aged over 50 years, and patients aged from 65 to 70 years corresponded to the largest group with the total fraction of 11.76% (12,229/103,945; Figure 3). The geolocation of patients was labeled as either intraprovince or interprovince. From 2015 to 2020, 98.70% (102,607/103,957) of teleconsultation encounters were labeled intraprovince, and the average annual increasing rate for this group was 0.41%; 1.30% (1350/103,957) of patients

XSL•FO
RenderX

were interprovince, with an average annual increasing rate of 21.27%. We noted that patients served by the telemedicine platform were mainly within the province (ie, Henan Province), but the service volume for interprovince cases increased significantly. There are 18 cities in Henan Province. When we zoom in on the group of patients within the province, a large proportion of them were from several cities including Nanyang, Pingdingshan, and Zhengzhou.

Figure 3. Distributions of gender and age of patients served.



Hospitals that applied for remote consultations contain all tiers (ie, tier 1, 2, and 3, where tier 3 hospitals have the highest ranks and the best medical resources). While 86.64% (90,069/103,957) of consultations were requested by tier 1 and tier 2 hospitals, only 13.36% (13,888/103,957) of cases were for tier 3 hospitals (Multimedia Appendix 2). Although the majority of applicants were tier 2 hospitals, the number of applications from tier 3 hospitals rapidly increased from 2015 to 2020 with an average annual rate of 47.12%. On the contrary, the changing rate for the group of tier 1 and tier 2 hospitals was –2.87%, which shows a dropping trend. As for the departments that applied for teleconsultation, internal medicine departments requested the majority (66,970/103,597, 64.42%) of cases, and 24.33% (25,295/103,597) of consultations were requested by surgery departments. Other disciplines included gynecology, pediatrics, and otolaryngology; the top three departments were neurology, respiratory medicine, and pediatric internal medicine. Table 1 shows the number of teleconsultations requested by each department and their percentage values for each year. Top 10 departments were mapped to 63.97% (66,276/103,597) of cases. Almost all (102,960/103,957, 99.04%) medical practitioners and experts from invited hospitals had senior professional titles (eg, professor).

Table 1. Top 10 departments applied for teleconsultation.

Department	2015, n (%)	2016, n (%)	2017, n (%)	2018, n (%)	2019, n (%)	2020, n (%)
Neurology	1834 (13.75)	2869 (14.80)	2985 (14.38)	2263 (11.62)	1935 (11.23)	1444 (10.49)
Respiratory medicine	1283 (9.62)	1898 (9.79)	2499 (12.04)	2643 (13.57)	2368 (13.75)	2122 (15.41)
Pediatric internal medicine	962 (7.21)	1199 (6.18)	1422 (6.85)	1570 (8.06)	1329 (7.72)	888 (6.45)
Oncology	764 (5.73)	1117 (5.76)	1291 (6.22)	1273 (6.53)	1083 (6.29)	955 (6.93)
Orthopedics	792 (5.94)	1206 (6.22)	1197 (5.77)	1064 (5.46)	1035 (6.01)	646 (4.69)
Cardiovascular medicine	850 (6.37)	885 (4.56)	1058 (5.10)	909 (4.67)	894 (5.19)	696 (5.05)
Neurosurgery	615 (4.61)	936 (4.83)	868 (4.18)	841 (4.32)	817 (4.74)	660 (4.79)
Gastroenterology	488 (3.66)	716 (3.69)	680 (3.28)	682 (3.50)	536 (3.11)	439 (3.19)
Hepatobiliary, pancreatic and liver transplantation	502 (3.76)	704 (3.63)	651 (3.14)	601 (3.09)	536 (3.11)	431 (3.13)
Gynecology	449 (3.37)	653 (3.37)	628 (3.03)	599 (3.07)	623 (3.62)	393 (2.85)

XSL•FO
RenderX

Teleconsultation Service Process

The waiting time between an application submitted by the applicant hospital and the start of the teleconsultation care reflects the efficiency of a case. Results from our analysis showed that the median value of waiting time was 24.27 hours, the minimum value was 11.00 minutes, and the maximum value was 7.00 days (Multimedia Appendix 2). From 2015 to 2020, the waiting time in different years appeared to be statistically different (rank sum test [H]=8309.00, *P*<.001). The median value for the year 2015 was 15.92 hours, while the value for the year 2020 was 24.27 hours. We observed an increasing trend in years for the waiting time. While the service capability in provincial hospitals remained almost unchanged, as our platform was promoted, more cases (especially those involving rare and intractable diseases) were diagnosed through teleconsultation encounters. As a consequence, the average waiting time increased.

The consultation duration between the start and end of teleconsultation reflects the quality and complexity of the service process to some extent. From 2015 to 2020, the median value was 17.00 minutes (Multimedia Appendix 2)—18.00 minutes in 2016 and 26.00 minutes in 2020. We noted that the consultation duration also increased among years statistically (H=3072.93, *P*<.001). The consultation duration was affected by many factors such as participating disciplines, condition of patients, and capability of medical practitioners. Therefore, we performed multivariable linear regression to analyze the influence of those key factors (eg, application content and patients) on consultation duration. In our regression analysis, consultation duration is the dependent variable. Independent variables contain consultation date, level (or tier) of the applicant hospital, gender of the patient, age of the patient, discipline (or department), title of the medical practitioner, and waiting time. Results show that independent variables such as year of consultation, applicant hospital level, and department were considered as significant inputs by the model (as shown in Table 2). Compared with the value in 2015, the average consultation duration increased in 2018, 2019, and 2020, with β values of 4.13, 3.08, and 8.16, respectively (*P*<.001). Consultation durations are usually longer when applicant institutes are tier 3 hospitals (β=1.69, *P*<.001) or patients have been referred (β=0.69, *P*=.01). Consultation durations for surgical departments are shorter than those for internal medicine departments (β=–1.00, *P*<.001).

Table 2. Results of multivariable linear regression analysis for impact factors of consultation durations.

Variable	β (95% CI)	SE	T score	*P* value
Constant	21.04 (19.70 to 22.37)	0.68	30.96	<.001
Year (ref[a]: 2015)				
2016	–4.00 (–4.36 to –3.64)	0.18	–21.93	<.001
2017	0.25 (–0.15 to 0.65)	0.21	1.23	.22
2018	4.31 (3.66 to 4.61)	0.24	17.10	<.001
2019	3.08 (2.58 to 3.59)	0.26	12.03	<.001
2020	8.16 (7.60 to 8.73)	0.29	28.40	<.001
Applicant hospital level (ref: tier 2 and below hospital)				
Tertiary	1.69 (1.29 to 2.09)	0.2	8.31	<.001
Waiting time	0.00 (0.00 to 0.00)	0	1.85	.06
Transfer treatment (ref: no)	0.69 (0.17 to 1.21)	0.27	2.6	.01
Consultation department (ref: internal medicine department)				
Surgery department	–1.00 (–1.30 to –0.71)	0.15	–6.63	<.001
Otolaryngology department	0.58 (–0.11 to 1.26)	0.35	1.64	.10
Gynecology & pediatrics departments	–0.92 (–1.47 to –0.36)	0.28	–3.25	.001
Medical technology department	0.28 (–0.49 to 1.05)	0.39	0.71	.48
Title of invited consultant (ref: attending doctor)				
Associate chief physician	–0.78 (–2.09 to 0.54)	0.67	–1.16	.25
Chief physician	–1.51 (–2.83 to –0.19)	0.67	–2.25	.03

[a]ref: reference.

Analysis of Withdrawal and Referral Cases

From 2015 to 2020, 5.09% (5270/103,957) of cases were withdrawn before the start of teleconsultation. Major reasons included the patient having been transferred (1209/5270, 22.94%) and the applicant medical practitioner requested a withdraw of consultation (1310/5270, 24.86%). This work investigates the characteristics of referral cases since 2017. After a teleconsultation, the invited medical experts may give referral advice accordingly. A referral decision is made by considering

XSL•FO
RenderX

a collection of factors such as severity of condition, physical distance (ie, within or outside Henan Province), and level of resident hospital. Therefore, we analyzed the impact of different factors on the referral cases. From 2017 to 2020, 15.27% (10,878/71,299) of cases received suggestions for referral from invited medical experts after consultations. We used binary logistic regression analysis to understand relations between those suggestions of referral and factors such as level of applicant hospitals, conditions of patients, and participated departments. From our results, we have reached 3 key observations. First, patients within Henan Province were more likely to receive suggestions of referral compared with those who were in other provinces (OR 2.17, *P*<.001). Second, older patients had a low referral rate (OR 1.00, *P*=.002). Third, compared with consultations hosted by the internal medicine department, those by surgery department, otolaryngology department, and gynecology and pediatrics departments had higher rates of referrals, with *P*<.001 and OR values of 2.94, 4.53, 3.51, respectively (Table 3).

Table 3. Results of binary logistic regression analysis for impact factors of referral cases.

Variable	β	SE	Wald	OR[a] (95% CI)	*P* value
Constant	–4.98	0.30	274.24	0.007	<.001
Region (ref [b]: other provinces)					
Henan province	0.77	0.22	12.34	2.17 (1.41-3.33)	<.001
Applicant hospital level (ref: tier 2 and below hospital)					
Tertiary	0.35	0.06	39.94	1.41 (1.27-1.57)	<.001
Age	–0.003	0.001	9.58	1.00 (1.00-1.00)	.002
Consultation department (ref: internal medicine department)					
Surgery department	1.08	0.20	29.88	2.94 (2.00-4.32)	<.001
Otolaryngology department	1.51	0.20	58.18	4.53 (3.07-6.67)	<.001
Gynecology & pediatrics departments	1.26	0.22	33.22	3.51 (2.29-5.37)	<.001
Medical technology department	1.43	0.21	47.45	4.12 (2.79-6.30)	<.001
Title of invited consultant (ref: attending doctor)					
Associate chief physician	0.49	0.17	8.34	1.63 (1.17-2.27)	.004
Chief physician	0.07	0.04	3.25	1.07 (1.00-1.16)	.07
Consultation time	0.02	0.001	256.81	1.02 (1.02-1.02)	<.001

[a]OR: odds ratio.

[b]ref: reference.

Attitudes Toward Telemedicine From Medical Practitioners and Patients

From our investigations into attitudes toward telemedicine from medical practitioners, we have observed that the majority of them held positive views; 68.4% (154/225) of participating experts believed that telemedicine is of great help in increasing the service quality of medical practitioners and 59.1% (133/225) believed that telemedicine is of great help in reducing financial burdens on patients. Considering the level of satisfaction toward consultation processes, 60.4% (136/225) of medical practitioners were very satisfied, and 76.9% (173/225) would recommend telemedicine services for their patients (Table 4). Patients also showed positive attitudes toward the telemedicine services they have received, with 60.0% (135/225) saying telemedicine is helpful for treatment of their disease, and 62.2% (140/225) agreeing that remote consultations provide convenience. In addition, 53.8% (121/225) were very satisfied with telemedicine services, and 59.1% (133/225) would suggest telemedicine to other patients (Table 5).

XSL•FO
RenderX

Table 4. Attitudes toward telemedicine from medical practitioners.

Item	Response, n (%)
Do you believe that telemedicine is helpful in improving the service quality of medical practitioners?	
Very helpful	154 (68.4)
Helpful	70 (31.1)
Not helpful	1 (0.4)
Do you believe that telemedicine can help to reduce the financial burden on patients?	
Very helpful	133 (59.1)
Helpful	85 (37.8)
Not helpful	7 (3.1)
What is your level of satisfaction with the service process of telemedicine?	
Very satisfied	136 (60.4)
Satisfied	88 (39.1)
Not satisfied	1 (0.4)
Would you like to participate in telemedicine services in the long term?	
Very likely	177 (78.7)
Likely	45 (20.0)
Not likely	3 (1.3)
Would you recommend telemedicine services for patients?	
Very likely	173 (76.9)
Likely	49 (21.8)
Not likely	3 (1.3)

Table 5. Attitudes toward telemedicine from patients.

Item	Response, n (%)
Do you believe that telemedicine is helpful for your treatments?	
Very helpful	135 (60.0)
Helpful	89 (39.6)
Not helpful	1 (0.4)
Do you believe that telemedicine reduces the financial cost of seeking medical treatment?	
Yes	215 (95.6)
No	2 (0.9)
Not sure	8 (3.6)
Does telemedicine provide convenience?	
Very convenient	140 (62.2)
Convenient	83 (36.9)
Not convenient	2 (0.9)
Are you satisfied with the service process of telemedicine?	
Very likely	133 (59.1)
Likely	91 (40.4)
Not likely	1 (0.4)

XSL•FO
RenderX

Discussion

Principal Findings

In recent years, telemedicine has been promoted rapidly in China. Many governmental policies and documents requesting the promotion of telemedicine to townships and rural areas have been issued. A large scope of services such as teleconsultation, telepathology, tele-electrocardiogram, and telediagnosis of medical images are widely offered with the business-to-business service mode. The current progress in telemedicine has greatly improved the service quality of primary medical institutions and brought convenience to patients. We note that further improvements in consultation efficiency and optimization of the consultation process are yet to be addressed. In this paper, we investigated the first and largest regional and comprehensive telemedicine platform deployed in China that changed legacy medical consultation methods (ie, via telephone and communication software) by realizing telemedicine services through platform technologies. It is a representative milestone toward the promotion of telemedicine in China. The platform that is supported by the largest hospital in Henan Province offers teleconsultation services to medical institutions at all levels within and outside the province. Since 2015, the service volume of this regional platform showed an increasing pattern followed by a slowly decreasing trend. With the adoption of telemedicine services by more medical institutions, municipal and county-level hospitals can host remote consultations for primary medical institutions. Therefore, teleconsultation cases for common diseases are offloaded to municipal or county-level hospitals that are more responsive than large institutions. During the pandemic of COVID-19, the patients with normal (or chronic) conditions preferred to defer their visits to hospitals, and the number of hospitalized patients at all levels of institutions decreased significantly. Thus, we saw a decreasing pattern of service volume for this provincial platform from 2018 to 2020. We note that the frequent connections between medical institutions at adjacent levels is one objective of the hierarchical medical system in China.

From 2015 to 2020, 103,957 teleconsultations were conducted through this platform. The majority of patients were from Henan Province with a much smaller percentage of interprovince patients, which is consistent with the results of existing studies [29]. Patients participating in remote consultations were mostly aged over 50 years, with higher risks in having diseases [30,31]. Applicant institutions were mainly tier 1 or 2 hospitals, defined as primary hospitals with limited medical capability, especially in the treatment of severe, intractable, and rare diseases [32]. Therefore, they need supports from large hospitals through the telemedicine platform. Our study also revealed the popular disciplines involved in teleconsultation encounters, including neurology, respiratory medicine, oncology, and cardiovascular medicine. As we know, the spectrum of human diseases has changed with the development of society. Chronic diseases such as stroke, malignant tumor, chronic obstructive pulmonary disease, heart disease, and circulatory disease have become critical problems for the community [33-35]. As a result, the number of patients with those issues keeps increasing [36], leading to high demands for medical consultations in related disciplines. To guarantee service quality, it is suggested by most provincial governments that the invited experts for medical consultations should have senior professional titles. In our study, we found that 99.04% of invited doctors met the requirements, which is much higher than the percentage of legacy face-to-face consultations. Only 27.5% of doctors in primary medical institutions (eg, township hospitals and village clinics) have proper professional titles, which is one of the reasons that they could not provide high-quality medical services [37,38]. As one motivation for the rapid development of telemedicine, patients of primary hospitals could also receive better diagnoses and treatments from experienced experts with senior titles.

Due to the large population of China, there is a high demand for medical resources. It is quite common for large hospitals that provide high-quality medical services to have an excessive number of patients, long waiting times, and degraded efficiency of medical treatments [39]. However, with the support from the telemedicine platform, the average waiting time after application submitted by applicant hospitals is 1388.00 minutes (ie, 23.13 hours), indicating that the majority of cases could be processed within a day, effectively increasing the efficiency of medical services. Some studies have reported that the average waiting time for the return of remote pathology results exceeds one day [40], and the waiting time for remote pathology consultation results in some countries exceeds 4 days [41]. In contrast, the waiting time for remote consultations supported by our platform is relatively short. The average consultation duration was 17.00 minutes. As discovered by our multivariable analysis, the consultation duration increases with time. The consultation quality increases with the advances of communication and digital conferencing technologies [42], which make it convenient for participants to have a more comprehensive and in-depth discussion. The consultation duration is longer for patients who received referrals as their conditions are rarer and more complicated.

After teleconsultation services, the invited medical experts may make referral suggestions. Compared with intraprovince patients, we found that interprovince patients are more likely to receive referral suggestions. It is mainly because that interprovince referrals are quite inconvenient due to long physical distances, and the government encourages patients to receive medical treatments within their home provinces. Elderly patients with limited mobility, for example, have a low referral rate. Since the quality of surgical operations is usually better in large hospitals, a higher referral rate is observed for cases in surgery departments.

As we found in our study, the majority of medical practitioners believe that telemedicine improves service qualities and they would like to participate in teleconsultation services, consistent with existing research findings [43,44]. The promotion of telemedicine is beneficial for doctors on both sides. Through remote consultations, the invited experts have more opportunities to gain experiences on intractable and rare diseases, while doctors from the applicant hospitals learn from their counterparts at large hospitals. Prior research works have shown that the overall satisfaction of doctors in traditional medical service processes is moderate, and the personal satisfaction of doctors is quite low [45,46]. However, we

XSL•FO
RenderX

illustrated that 60.4% of medical practitioners experienced a high level of satisfaction with the process of telemedicine services, indicating its popularity among doctors.

Most doctors and patients agreed that the deployment of telemedicine could save financial costs, which is consistent with the views of existing research [47]. Some researchers found that telemedicine could help patients save US $1000, on average [48]. Higher level hospitals provide lower level medical institutions with remote consultations so that patients from less capable hospitals also receive advice from experts with senior titles, which is very beneficial for the patients in rural areas. We found that 60.0% of patients believe telemedicine is helpful in the treatment of their medical conditions, and 53.8% are very satisfied with the services they have received.

Limitations

The paper analyzes the process of remote consultation and provides an important reference for improving service efficiency. However, the types of telemedicine services are diverse, and research on the other service types such as remote education and remote nursing needs to be further developed. In addition, the data used in this paper were collected from a regional telemedicine platform, which cannot fully represent the development of telemedicine in China, and the overall application of telemedicine services needs further exploration.

Conclusions

In this paper, we describe the architecture and functionality of the first regional telemedicine center in China built using platform technologies. We collected and analyzed teleconsultation care services conducted through the platform from 2015 to 2020. Our work reveals the growth trend of service volume, reveals that the majority of applicant institutions are tier 2 hospitals, and shows that middle-aged and elderly patients make up the largest age group. In China, the promotion and deployment of telemedicine services is occurring at a rapid speed, which meets the urgent demand of medical institutions with less capability and the medical need of aged patients with less mobility. The efficiency of telemedicine is increasing with the advances of information technologies. Both medical practitioners and patients have high levels of satisfaction. It is believed that the deployment of a telemedicine platform not only increases the efficiency of medical consultations but also reduces the financial burdens of patients; thus, telemedicine is worthy of further promotion.

Acknowledgments

This research was supported by grant 2017YFC0909900 from the National Key Research and Development Program of China, grant 20IRTSTHN028 from the Innovation Research Team of Higher Education in Henan Province, and grant LHGJ20200349 from the Joint Co-Construction Program of Medical Science and Technology of Henan Province. The funders of the study played no role in the design, data collection, or interpretation of this work. The views expressed in the article are those of the authors and do not necessarily reflect the position of the funding bodies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of the hospitals connected to the regional platform.
[PDF File (Adobe PDF File), 314 KB - jmir_v23i7e28009_app1.pdf]

Multimedia Appendix 2
Basic information in the teleconsultations.
[PDF File (Adobe PDF File), 222 KB - jmir_v23i7e28009_app2.pdf]

References

1. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet 2008 Nov 15;372(9651):1774-1781. [doi: 10.1016/S0140-6736(08)61363-X] [Medline: 18930528]
2. Sun J, Luo H. Evaluation on equality and efficiency of health resources allocation and health services utilization in China. Int J Equity Health 2017 Jul 14;16(1):127 [FREE Full text] [doi: 10.1186/s12939-017-0614-y] [Medline: 28709422]
3. Li X, Krumholz HM, Yip W, Cheng KK, De Maeseneer J, Meng Q, et al. Quality of primary health care in China: challenges and recommendations. Lancet 2020 Jun 06;395(10239):1802-1812 [FREE Full text] [doi: 10.1016/S0140-6736(20)30122-7] [Medline: 32505251]
4. Li H, Liu K, Gu J, Zhang Y, Qiao Y, Sun X. The development and impact of primary health care in China from 1949 to 2015: a focused review. Int J Health Plann Manage 2017 Jul;32(3):339-350. [doi: 10.1002/hpm.2435] [Medline: 28670754]
5. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet 2019 Sep 28;394(10204):1192-1204. [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
6. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: a new health care mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]

XSL•FO
RenderX

7. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
8. He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, et al. Using an internet-based hospital to address maldistribution of health care resources in rural areas of Guangdong province, China: retrospective and descriptive study. JMIR Med Inform 2018 Dec 21;6(4):e51 [FREE Full text] [doi: 10.2196/medinform.9495] [Medline: 30578195]
9. Wang Z, Gu H. A review of telemedicine in China. J Telemed Telecare 2009;15(1):23-27. [doi: 10.1258/jtt.2008.080508] [Medline: 19139216]
10. Cai H, Wang H, Guo T, Bao G. Application of telemedicine in Gansu province of China. PLoS One 2016 Jun 22;11(6):e0158026 [FREE Full text] [doi: 10.1371/journal.pone.0158026] [Medline: 27332894]
11. Wang T, Li J, Zhu C, Hong Z, An D, Yang H, et al. Assessment of utilization and cost-effectiveness of telemedicine program in western regions of China: a 12-year study of 249 hospitals across 112 cities. Telemed J E Health 2016 Nov;22(11):909-920. [doi: 10.1089/tmj.2015.0213] [Medline: 27314300]
12. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18. [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
13. Song X, Liu X, Wang C. The role of telemedicine during the COVID-19 epidemic in China-experience from Shandong province. Crit Care 2020 Apr 28;24(1):178 [FREE Full text] [doi: 10.1186/s13054-020-02884-9] [Medline: 32345359]
14. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
15. Weinstein RS, Krupinski EA, Doarn CR. Clinical examination component of telemedicine, telehealth, mhealth, and connected health medical practices. Med Clin North Am 2018 May;102(3):533-544. [doi: 10.1016/j.mcna.2018.01.002] [Medline: 29650074]
16. Zheng Y, Lin Y, Cui Y. Teledermatology in China: history, current status, and the next step. J Investig Dermatol Symp Proc 2018 Dec;19(2):S71-S73 [FREE Full text] [doi: 10.1016/j.jisp.2018.09.003] [Medline: 30471754]
17. Zangara G, Valentino F, Spinelli G, Valenza M, Marcheggiani A, Di Blasi F. An Albanian open source telemedicine platform. Telemed J E Health 2014 Jul;20(7):673-677. [doi: 10.1089/tmj.2013.0239] [Medline: 24906118]
18. Crespo LF, Morillo DS. AMICA telemedicine platform: a design for management of elderly people with COPD. 2009 Presented at: IEEE 9th International Conference on Information Technology and Applications in Biomedicine; 2009; Larnaka. [doi: 10.1109/itab.2009.5394296]
19. Beer J, Abrouk M, Kirsner R. Telemedicine platforms used in academic dermatology during the COVID-19 pandemic: implications for adaptation and usage. J Drugs Dermatol 2020 Aug 01;19(8):797-798. [doi: 10.36849/JDD.2020.401] [Medline: 32845113]
20. Burke BL, Hall RW. Telemedicine: pediatric applications. Pediatrics 2015 Jul;136(1):e293-e308 [FREE Full text] [doi: 10.1542/peds.2015-1517] [Medline: 26122813]
21. Maia MR, Castela E, Pires A, Lapão LV. How to develop a sustainable telemedicine service? A pediatric telecardiology service 20 years on—an exploratory study. BMC Health Serv Res 2019 Sep 23;19(1):681 [FREE Full text] [doi: 10.1186/s12913-019-4511-5] [Medline: 31547824]
22. Trettel A, Eissing L, Augustin M. Telemedicine in dermatology: findings and experiences worldwide—a systematic literature review. J Eur Acad Dermatol Venereol 2018 Feb;32(2):215-224. [doi: 10.1111/jdv.14341] [Medline: 28516492]
23. Kahn EN, La Marca F, Mazzola CA. Neurosurgery and telemedicine in the United States: assessment of the risks and opportunities. World Neurosurg 2016 May;89:133-138. [doi: 10.1016/j.wneu.2016.01.075] [Medline: 26852710]
24. Xu T, Pujara S, Sutton S, Rhee M. Telemedicine in the management of type 1 diabetes. Prev Chronic Dis 2018 Jan 25;15:E13 [FREE Full text] [doi: 10.5888/pcd15.170168] [Medline: 29369757]
25. McWilliams T, Hendricks J, Twigg D, Wood F, Giles M. Telehealth for paediatric burn patients in rural areas: a retrospective audit of activity and cost savings. Burns 2016 Nov;42(7):1487-1493. [doi: 10.1016/j.burns.2016.03.001] [Medline: 27575678]
26. Hasselberg MJ. The digital revolution in behavioral health. J Am Psychiatr Nurses Assoc 2020;26(1):102-111. [doi: 10.1177/1078390319879750] [Medline: 31587624]
27. Martinez KA, Rood M, Jhangiani N, Kou L, Rose S, Boissy A, et al. Patterns of use and correlates of patient satisfaction with a large nationwide direct to consumer telemedicine service. J Gen Intern Med 2018 Oct;33(10):1768-1773 [FREE Full text] [doi: 10.1007/s11606-018-4621-5] [Medline: 30112737]
28. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 23;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
29. Wang X. Present situation and existing problems of telemedicine in Shandong Province In Chinese. J Shandong Univ (Health Sci) 2016 Feb;54(2):90-94. [doi: 10.6040/j.issn.1671-7554.0.2015.859]
30. Maresova P, Javanmardi E, Barakovic S, Barakovic Husic J, Tomsone S, Krejcar O, et al. Consequences of chronic diseases and other limitations associated with old age—a scoping review. BMC Public Health 2019 Nov 01;19(1):1431 [FREE Full text] [doi: 10.1186/s12889-019-7762-5] [Medline: 31675997]

XSL•FO
RenderX

31. Lyu J, Zhang W, Li W, Wang S, Zhang J. Epidemic of chronic diseases and the related healthy lifestyle interventions in rural areas of Shandong Province, China. BMC Public Health 2020 May 01;20(1):606 [FREE Full text] [doi: 10.1186/s12889-020-08729-y] [Medline: 32357867]
32. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int J Equity Health 2017 Mar 03;16(1):42 [FREE Full text] [doi: 10.1186/s12939-017-0543-9] [Medline: 28253876]
33. GBD 2016 DALYsHALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
34. Yang G, Wang Y, Zeng Y, Gao GF, Liang X, Zhou M, et al. Rapid health transition in China, 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet 2013 Jun 8;381(9882):1987-2015. [doi: 10.1016/S0140-6736(13)61097-1] [Medline: 23746901]
35. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
36. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
37. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
38. Zhang Q, Chen J, Yang M, Pan J, Li X, Yue L, et al. Current status and job satisfaction of village doctors in western China. Medicine (Baltimore) 2019 Aug;98(32):e16693 [FREE Full text] [doi: 10.1097/MD.0000000000016693] [Medline: 31393371]
39. Li Y, Gong W, Kong X, Mueller O, Lu G. Factors associated with outpatient satisfaction in tertiary hospitals in China: a systematic review. Int J Environ Res Public Health 2020 Sep 27;17(19):1 [FREE Full text] [doi: 10.3390/ijerph17197070] [Medline: 32992600]
40. Zhou C, Rahemtulla A, Hua LT, Shi H. Telepathology consultation in China using whole slide image and an internet based platform. Diagn Pathol 2013 Sep 30;8(S1):1. [doi: 10.1186/1746-1596-8-s1-s10]
41. Völker H, Müller-Hermelink H, Stüfe A, Strehl A, Pötzl L, Stauch G. [Ten years of telepathology for a mission hospital in Tanzania]. Pathologe 2019 Sep;40(5):519-526. [doi: 10.1007/s00292-019-0641-0] [Medline: 31338564]
42. Parimbelli E, Bottalico B, Losiouk E, Tomasi M, Santosuosso A, Lanzola G, et al. Trusting telemedicine: a discussion on risks, safety, legal implications and liability of involved stakeholders. Int J Med Inform 2018 Apr;112:90-98. [doi: 10.1016/j.ijmedinf.2018.01.012] [Medline: 29500027]
43. Velázquez M, Pacheco A, Silva M, Sosa D. [Evaluation of the teleconsultation process from the perspective of the provider (Oaxaca Telehealth Program, Mexico)]. Rev Panam Salud Publica 2017 May 25;41:e22 [FREE Full text] [Medline: 28591329]
44. Wichman CL, Laszewski A, Doering JJ, Borchardt S. Feasibility of model adaptations and implementation of a perinatal psychiatric teleconsultation program. Gen Hosp Psychiatry 2019;59:51-57. [doi: 10.1016/j.genhosppsych.2019.05.007] [Medline: 31163299]
45. Deng S, Yang N, Li S, Wang W, Yan H, Li H. Doctors' job satisfaction and its relationships with doctor-patient relationship and work-family conflict in China: a structural equation modeling. Inquiry 2018;55:46958018790831 [FREE Full text] [doi: 10.1177/0046958018790831] [Medline: 30371128]
46. Domagała A, Peña-Sánchez JN, Dubas-Jakóbczyk K. Satisfaction of physicians working in Polish hospitals—a cross-sectional study. Int J Environ Res Public Health 2018 Nov 25;15(12):2640 [FREE Full text] [doi: 10.3390/ijerph15122640] [Medline: 30477273]
47. Kruse CS, Lee K, Watson JB, Lobo LG, Stoppelmoor AG, Oyibo SE. Measures of effectiveness, efficiency, and quality of telemedicine in the management of alcohol abuse, addiction, and rehabilitation: systematic review. J Med Internet Res 2020 Jan 31;22(1):e13252 [FREE Full text] [doi: 10.2196/13252] [Medline: 32012048]
48. Zhai Y, Gao J, Chen B, Shi J, Wang L, He X, et al. Design and application of a telemedicine system jointly driven by videoconferencing and data exchange: practical experience from Henan Province, China. Telemed J E Health 2020 Jan 01;26(1):89-100. [doi: 10.1089/tmj.2018.0240] [Medline: 31050599]

XSL•FO
RenderX

Edited by R Kukafka; submitted 17.02.21; peer-reviewed by C Doarn, C Huang, W Zhang, M Usher; comments to author 20.03.21; revised version received 08.04.21; accepted 06.05.21; published 12.07.21.

Please cite as:
Cui F, He X, Zhai Y, Lyu M, Shi J, Sun D, Jiang S, Li C, Zhao J
Application of Telemedicine Services Based on a Regional Telemedicine Platform in China From 2014 to 2020: Longitudinal Trend Analysis
J Med Internet Res 2021;23(7):e28009
URL: https://www.jmir.org/2021/7/e28009
doi:10.2196/28009
PMID:34255686

©Fangfang Cui, Xianying He, Yunkai Zhai, Minzhao Lyu, Jinming Shi, Dongxu Sun, Shuai Jiang, Chenchen Li, Jie Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinical Outcomes of Asynchronous Versus Synchronous Telepsychiatry in Primary Care: Randomized Controlled Trial

Peter M Yellowlees[1], MBBS, MD; Michelle Burke Parish[1], MA, PhD; Alvaro D Gonzalez[1], MA; Steven R Chan[2,3], MBA, MD; Donald M Hilty[4], MD; Byung-Kwang Yoo[5], PhD, MD; J Paul Leigh[5], PhD; Robert M McCarron[6], DO; Lorin M Scher[1], MD; Andres F Sciolla[1], MD; Jay Shore[7], MD; Glen Xiong[1], MD; Katherine M Soltero[8], LCSW; Alice Fisher[5], BA; Jeffrey R Fine[5], MPH; Jennifer Bannister[1], EdM; Ana-Maria Iosif[5], PhD

[1]Department of Psychiatry and Behavioral Sciences, University of California Davis, Sacramento, CA, United States

[2]Stanford University School of Medicine, Stanford, CA, United States

[3]Veterans Administration Palo Alto Healthcare System, Palo Alto, CA, United States

[4]Northern California Veterans Administration, Mather, CA, United States

[5]Department of Public Health Sciences, University of California Davis, Davis, CA, United States

[6]University of California Irvine, Irvine, CA, United States

[7]University of Colorado Anschutz Medical Campus, Denver, CO, United States

[8]CommuniCare Health Centers, West Sacramento, Sacramento, CA, United States

Corresponding Author:
Peter M Yellowlees, MBBS, MD
Department of Psychiatry and Behavioral Sciences
University of California Davis
Sacramento, CA
United States
Phone: 1 916 734 0849
Email: pmyellowlees@ucdavis.edu

Abstract

Background: Asynchronous telepsychiatry (ATP; delayed-time) consultations are a novel form of psychiatric consultation in primary care settings. Longitudinal studies comparing clinical outcomes for ATP with synchronous telepsychiatry (STP) are lacking.

Objective: This study aims to determine the effectiveness of ATP in improving clinical outcomes in English- and Spanish-speaking primary care patients compared with STP, the telepsychiatry *usual care* method.

Methods: Overall, 36 primary care physicians from 3 primary care clinics referred a heterogeneous sample of 401 treatment-seeking adult patients with nonurgent psychiatric disorders. A total of 184 (94 ATP and 90 STP) English- and Spanish-speaking participants (36/184, 19.6% Hispanic) were enrolled and randomized, and 160 (80 ATP and 80 STP) of them completed baseline evaluations. Patients were treated by their primary care physicians using a collaborative care model in consultation with the University of California Davis Health telepsychiatrists, who consulted with patients every 6 months for up to 2 years using ATP or STP. Primary outcomes (the clinician-rated Clinical Global Impressions [CGI] scale and the Global Assessment of Functioning [GAF]) and secondary outcomes (patients' self-reported physical and mental health and depression) outcomes were assessed every 6 months.

Results: For clinician-rated primary outcomes, ATP did not promote greater improvement than STP at 6-month follow-up (ATP vs STP, adjusted difference in follow-up at 6 months vs baseline differences for CGI: 0.2, 95% CI −0.2 to 0.6; *P*=.28; and GAF: −0.6, 95% CI −3.1 to 1.9; *P*=.66) or 12-month follow-up (ATP vs STP, adjusted difference in follow-up at 12 months vs baseline differences for CGI: 0.4, 95% CI −0.04 to 0.8; *P*=.07; and GAF: −0.5, 95% CI −3.3 to 2.2; *P*=.70), but patients in both arms had statistically and clinically significant improvements in both outcomes. There were no significant differences in improvement from baseline between ATP and STP on any patient self-reported ratings at any follow-up (all *P* values were between .17 and .96). Dropout rates were higher than predicted but similar between the 2 arms. Of those with baseline visits, 46.8% (75/160) did not have a follow-up at 1 year, and 72.7% (107/147) did not have a follow-up at 2 years. No serious adverse events were associated with the intervention.

XSL•FO
RenderX

Conclusions: This is the first longitudinal study to demonstrate that ATP can improve clinical outcomes in English- and Spanish-speaking primary care patients. Although we did not find evidence that ATP is superior to STP in improving clinical outcomes, it is potentially a key part of stepped mental health interventions available in primary care. ATP presents a possible solution to the workforce shortage of psychiatrists and a strategy for improving existing systems of care.

Trial Registration: ClinicalTrials.gov NCT02084979; https://clinicaltrials.gov/ct2/show/NCT02084979.

(J Med Internet Res 2021;23(7):e24047) doi:10.2196/24047

KEYWORDS

asynchronous telepsychiatry; synchronous telepsychiatry; psychiatrist; primary care physician; psychiatric consultation; Spanish-speaking; collaborative care; workforce; depression; telehealth

Introduction

Background

Telepsychiatry, in the form of videoconferencing, is now an important tool in behavioral health care. For more than 30 years, synchronous telepsychiatry (STP), where consultations are performed in real time and are interactive, has increased access to care, making psychiatric experts available in areas with provider shortages. Research has demonstrated high levels of patient satisfaction and similar clinical outcomes to traditional in-person care for many disorders, including depression and anxiety [1,2]. Telemedicine utilization across all disciplines had already been anticipated to grow exponentially to a 130 billion dollar industry by 2025 [3], before the use of telepsychiatry dramatically increased during the COVID-19 pandemic. Suddenly, telepsychiatry has become a core health care tool [4] for most psychiatrists in the United States. Many clinics rapidly converted to telepsychiatry, with a number of them describing the experience and the changes required including the move to in-home consultations or virtual house calls. For example, the large University of California Davis (UCD) behavioral health outpatient clinic saw a successful conversion from approximately 97% in-person consultations to 100% virtual consultations in 3 days [4]. A survey conducted by the American Psychiatric Association during the COVID-19 pandemic found that by June 2020, 85% of 500 surveyed American psychiatrists were using telepsychiatry with more than 75% of their patients, compared with about 3% before COVID-19 [5]. The latest available national telehealth statistics derived from 60 contributing private insurers to the Fair Health database [6] as of December 2020 showed an increase of 2816% in telehealth consultations in all disciplines compared with December 2019, and these comprised 6.5% of all consultations nationally in their database, with 47% of the patients being seen primarily for mental health reasons [6]. The National Center for Health Statistics [7] reported a total of 883 million outpatient consultations nationally in 2018. Projecting from the insurance statistics [6], we can assume that about 3% of these visits in 2020 were telepsychiatry visits (by video or phone), an approximate total of 26 million such visits. There seems no doubt that STP has now become a major delivery component of mental health services in the United States.

Asynchronous Care

Despite this success, with STP being the current standard telepsychiatry practice [8,9], administrative and technical challenges exist, especially around scheduling of telepsychiatrists and patients [10,11], and STP itself is simply a virtual extension of in-person care that cannot be scaled to enable a physician to see more patients. Asynchronous care makes use of a completely virtual care model with the transmission of clinical information via web applications for review by a specialist at a later time [12,13] and has the potential to scale and enable psychiatrists to be involved in the treatment of more patients than with STP. Asynchronous care can also reduce the impact of poor bandwidth and connectivity issues seen with STP, providing potentially better access to more diverse patient populations. In recent years, asynchronous technologies have become more widespread in health care settings, utilized in patient portal email and messaging, in-app messaging, specialty patient-to-provider mobile apps, and forwarded interview videos [14]. Asynchronous technologies are commonly used in radiology, dermatology, ophthalmology, cardiology, and pathology [12,15-22] and are expanding into mental health care where they may be at least a partial solution to address the psychiatrist workforce shortage and reduce access barriers for patients [14]. Studies indicate that up to 10% of those who reported experiencing mental health difficulties are able to utilize available asynchronous resources [23], which can improve self-management and access. Positive patient outcomes have been demonstrated with e-coaching for depression [24], mobile-based asynchronous text-messaging therapy with a licensed therapist [25], integrated asynchronous virtual care platform [26], and asynchronous telehealth [27]. In addition to improved patient outcomes, such services have been found to increase access and care quality and reduce overutilization and costs of care [26].

Asynchronous telepsychiatry (ATP) is a more data-rich form of the traditional medical or psychiatric *curbside consultation*. In ATP, a trained interviewer conducts and records a semistructured patient interview, which is combined with other available clinical data, such as electronic medical records (EMRs). This recorded video consult and information from the EMR is made accessible to a telepsychiatrist who reviews it before providing an opinion on the patient's diagnosis and treatment options. The process, including consultation templates, has been fully described in previous publications [8,28-33]. Early pilot studies provide evidence that ATP has similar diagnostic accuracy to STP in English- and Spanish-speaking patients, that it is a feasible consultation modality in primary care patients and with patients cared for in skilled nursing

XSL•FO
RenderX

facilities [28-35], and that it may also be less costly to implement.

Objectives

This paper describes the first large study conducted to determine the effectiveness of ATP in improving clinical outcomes, as compared with STP, the current gold standard telepsychiatry *usual care* treatment method. We hypothesized that, compared with participants in the STP arm, participants in the ATP arm would show better clinical trajectories throughout treatment, as measured by greater improvements in clinician and patient self-reported ratings of global functioning, health outcomes, and depression.

Methods

Study Design and Setting

The study was a randomized controlled clinical trial conducted at 3 community-based primary care clinics in the Sacramento area, with patient recruitment occurring between March 2014 and September 2018. A data and safety monitoring board (DSMB), consisting of 2 independent physicians and 1 statistician, periodically reviewed and evaluated the accumulated study data for participant safety, study conduct, and progress. In a DSMB review in early 2018, it was noted that the dropout rate at 24 months was higher than anticipated. Thus, the DSMB recommended that the primary end point be at 12 months and advised subsequent enrollment be limited to the 12-month follow-up. The 12-month follow-up was the primary analysis of interest, and the institutional review board (IRB) documentation was modified in April 2018. The last 18 patients were enrolled for 1 year.

Participants

We recruited 36 primary care physicians (PCPs) as referring providers. We placed an alert in the electronic medical record system to remind the PCPs about the trial. Patients learned about the study through their PCP or from advertisements at the referring clinics. All participants were aged 18 years or older, able to give written informed consent, and were referred by PCPs as having one or more nonurgent mental health diagnoses, mainly mood disorders, anxiety disorders, or substance and alcohol use disorders. Many patients had comorbid conditions and multiple diagnoses. We attempted to overenroll Spanish-speaking patients and included a Federally Qualified Health Center that primarily treats Spanish-speaking patients as one of our referring clinics. The study protocol was approved by the UCD IRB; written informed consent was obtained from both patients and the referring PCPs before participation.

Potential participants completed a multistep screening and enrollment process. This consisted of a semistructured phone interview as well as the Patient Health Questionnaire-9 (PHQ-9) [36] to screen for risk of suicidality, followed by an in-person assessment with the Structured Clinical Interview for the Diagnostic Statistical Manual (SCID) IV Diagnoses in English or Spanish [37]. The SCID established a primary axis I diagnosis, which was used for stratified randomization. Before enrolling the first participant, the study statistician created a stratified block randomization schedule for each study site. Within each site, patients were assigned 1:1 to the 2 intervention arms in random permuted blocks of size 4 generated for each SCID primary axis I categorization to reduce imbalance between arms.

Patients were recruited for 4 study years and were followed up between 1 and 2 years with ATP or STP evaluations every 6 months (up to 5 visits). Four UCD faculty psychiatrists provided consultations for both ATP and STP groups, with a bilingual psychiatrist seeing all the patients who could only speak Spanish or preferred to have their consultations in that language. All psychiatrists were fully trained to deliver both types of consultations. Diagnostic conclusions and treatment recommendations of the consulting psychiatrist for all patients were reported back to the PCP in the psychiatrists' notes in the EMR. The PCPs then implemented the recommendations at their own discretion and could also communicate further by secure messaging or phone with the psychiatrists if they wished. ATP interviewers in the trial were behavioral health clinicians with a master's degree or higher, and their training for this trial has already been described [38].

Intervention Arm

ATP assessments were conducted at 6-month intervals by an ATP-trained clinician who spoke the patient's primary language, either English or Spanish [38]. This interview was video-recorded using Health Insurance Portability and Accountability Act–compliant security systems and protocols. For each ATP assessment, the clinician updated a standardized electronic form to capture notes about clinically relevant or important materials observed during the interview. These notes were usually completed the day of the ATP interview so that study psychiatrists had rapid access to the entire interview video, the clinician's interview notes, and previous medical and sometimes psychiatric assessments of the patient already recorded in their EMR. After each ATP visit, the psychiatrist provided the patient's PCP with a written assessment and psychiatric treatment plan. The PCP also had continued access to this psychiatrist by phone or email between the study consultations for up to 2 years [8,38].

Control Arm

The clinical workflow process for the STP arm was similar to that of the ATP arm, except that ATP-recorded assessments were replaced by live real-time STP assessments conducted by a psychiatrist who spoke the patient's preferred language, either English or Spanish. After the STP consultation, the psychiatrist provided the patient's PCP with a written assessment and treatment plan in the patient's EMR and was available for future contact by phone or email as necessary.

A demographic questionnaire was administered at the baseline to collect sociodemographic information. Participants were clinically assessed in both study arms at 6-month intervals (baseline, 6 months, 12 months, 18 months, and 24 months), with the primary outcome measures completed by the treating psychiatrists. All other study questionnaires assessing self-reported outcomes were collected every 6 months by research assistants either by phone or via paper or electronic surveys, depending on participants' preferences. Participants

XSL•FO
RenderX

were compensated for each assessment visit with a US $25 gift card, an amount considered by the IRB to be noncoercive. The PCPs were not compensated.

Primary Outcomes

The primary outcomes were derived from the psychiatrist's report and included the Clinical Global Impressions (CGIs) scale [39], which focuses primarily on functional impairment, and the Global Assessment of Functioning (GAF) [40], which mainly measures symptom severity. The CGI is a 3-item, 7-point observer-rated scale that measures illness severity, global improvement or change, and therapeutic response. The CGI is considered a robust measure with established validity in inpatient [41], outpatient [42], and clinical trial settings [42]. The CGI severity of illness and improvement scales are commonly used in nondrug trial settings [39]. We used the CGI severity of illness scale scored from 1 (normal) to 7 (among the most extremely ill). The GAF is a widely used rating scale for assessing impairment among patients with psychiatric disorders. The GAF assesses the level of psychological, social, and occupational functioning on a scale of 1 to 100, with higher levels indicating better functioning [40].

Secondary Outcomes

Secondary outcomes focused on patient self-report and included the 12-Item Short-Form Health Survey Physical Health Summary (PHS-12) and 12-Item Short-Form Health Survey Mental Health Summary (MHS-12) [43] scores (both scored from 0-100, with higher scores indicating better health) and the PHQ-9 [44]. The PHQ-9 is a well-validated depression scale with scores derived as the sum of 9 items (each scored from 0 [not at all] to 3 [nearly every day]; scale range 0-27) based directly on the diagnostic criteria for major depressive disorder in the Diagnostic and Statistical Manual Fourth Edition [37].

Sample Size Calculation

The statistical power to assess the difference in improvement from baseline to the primary end point (12 months) between ATP and STP for the clinician and patient-reported measures was evaluated assuming that 100 patients would be randomized into each arm, an attrition rate of 25%, and that half an SD was the smallest difference that would be clinically meaningful. Assuming a type 1 error α=.05, a two-sided test, measurements at baseline, 6, and 12 months, correlations between repeated measures ranging from 0.30 to 0.60, and the proposed sample size of 75 patients in each arm (after 25% loss to follow-up), we anticipated having 83%-96% power to detect half an SD difference in improvement between the arms at 12 months. On the basis of published data, we estimated a residual SD of 1.5 for CGI [45] and 10-12 points for the mental (MHS-12) and physical health (PHS-12) subscales of the 12-Item Short-Form Health Survey [46]. Thus, under the above assumptions, this study was sufficiently powered to detect a difference in improvement between the arms at 12 months of 0.75 for CGI and 5-6 points on MHS-12 and PHS-12. The calculation assumes 75 patients per arm, but some of the patients lost to follow-up would have completed some evaluations and contributed data, and thus the power would be higher. In addition, the power would be higher if the correlation between repeated measures was greater than 0.6.

Statistical Analysis

Group differences in demographic and clinical characteristics were assessed using χ^2 test (or Fisher exact test) for categorical variables and the two-sample (two-tailed) *t* tests (or Wilcoxon two-sample test) for continuous variables, as appropriate.

All analyses were intention-to-treat, and patients were analyzed as randomized. Mixed-effects linear regression models [47] were used to characterize the longitudinal trajectories of primary and secondary outcomes and assess intervention effects. This approach explicitly accounts for multiple measurements per person, allows for unequally spaced and missing observations, and produces valid inferences under the assumption that data are missing at random. The primary end point was the 12-month follow-up, and all participants who had at least 1 follow-up clinician rating at 6 or 12 months were included in the primary analysis. For each outcome variable, we fit a model that included terms for the intervention arm (ATP or STP), time (baseline, 6 months, and 12 months), and the interaction between arm and time. Models for clinician-rated outcomes were adjusted for a composite variable whose levels captured all possible combinations of study sites, treating psychiatrists, and language of the interview. Models for patient self-reported outcomes were similarly adjusted for a composite variable whose levels captured all possible combinations of study site, person conducting the interview (ATP interviewer or STP psychiatrist), and language of the interview. We accounted for clustering using a random effect for the patient and, whenever possible, a random effect for the referring physician. The interaction terms allowed us to assess intervention effects, that is, adjusted differences in follow-up compared with baseline differences between ATP and STP. All contrasts were estimated with 95% CIs and tested with two-sided alternatives using P<.05 as a threshold for statistical significance. No adjustments were made for multiple comparisons. Secondary analyses to confirm the longitudinal pattern from the primary analyses were conducted using the data up to 24 months and included all participants with at least 1 follow-up visit. Sensitivity analyses controlling for baseline values were conducted to confirm the primary analysis results. All analyses were performed using PROC MIXED in SAS version 9.4 (SAS Institute Inc) [48].

Results

Overview

Of the 36 consented clinicians, 28 (78%) referred at least one patient. Figure 1 depicts the flow of patients from screening through the primary end point and the 12-month follow-up. Of the 401 patients assessed for eligibility, 184 (45.9%) were enrolled and randomized to the ATP (n=94) or STP (n=90) intervention. Of the 184 randomized participants, 18 (9.8%; 11 ATP and 7 STP) consented to the 12-month follow-up, and 24 (13%; 14 ATP and 10 STP) withdrew before the baseline visit. Reasons for withdrawal before baseline included insurance changes (n=2), decline to participate (n=7), and loss to follow-up (n=15). Multimedia Appendix 1, Table S1 shows the

XSL•FO
RenderX

demographic and clinical characteristics of the 160 participants who completed the baseline visit and the 24 who did not. The 2 groups were similar in terms of sociodemographic characteristics and depression symptoms, but participants who completed the baseline visit were more likely to be receiving current outpatient psychotherapy for a psychiatric condition (65/158, 41.1% vs 5/24, 21%; *P*=.06) and to be using psychotropic medication (130/157, 82.8% vs 12/24, 50%; *P*<.001) than those who did not complete baseline visits. Interestingly, only 1 of these 160 patients who completed a baseline visit was seeing an outpatient psychiatrist, with the rest being treated in primary care.

Figure 1. Participants flow through 12-month follow-up. ATP: asynchronous telepsychiatry; STP: synchronous telepsychiatry.

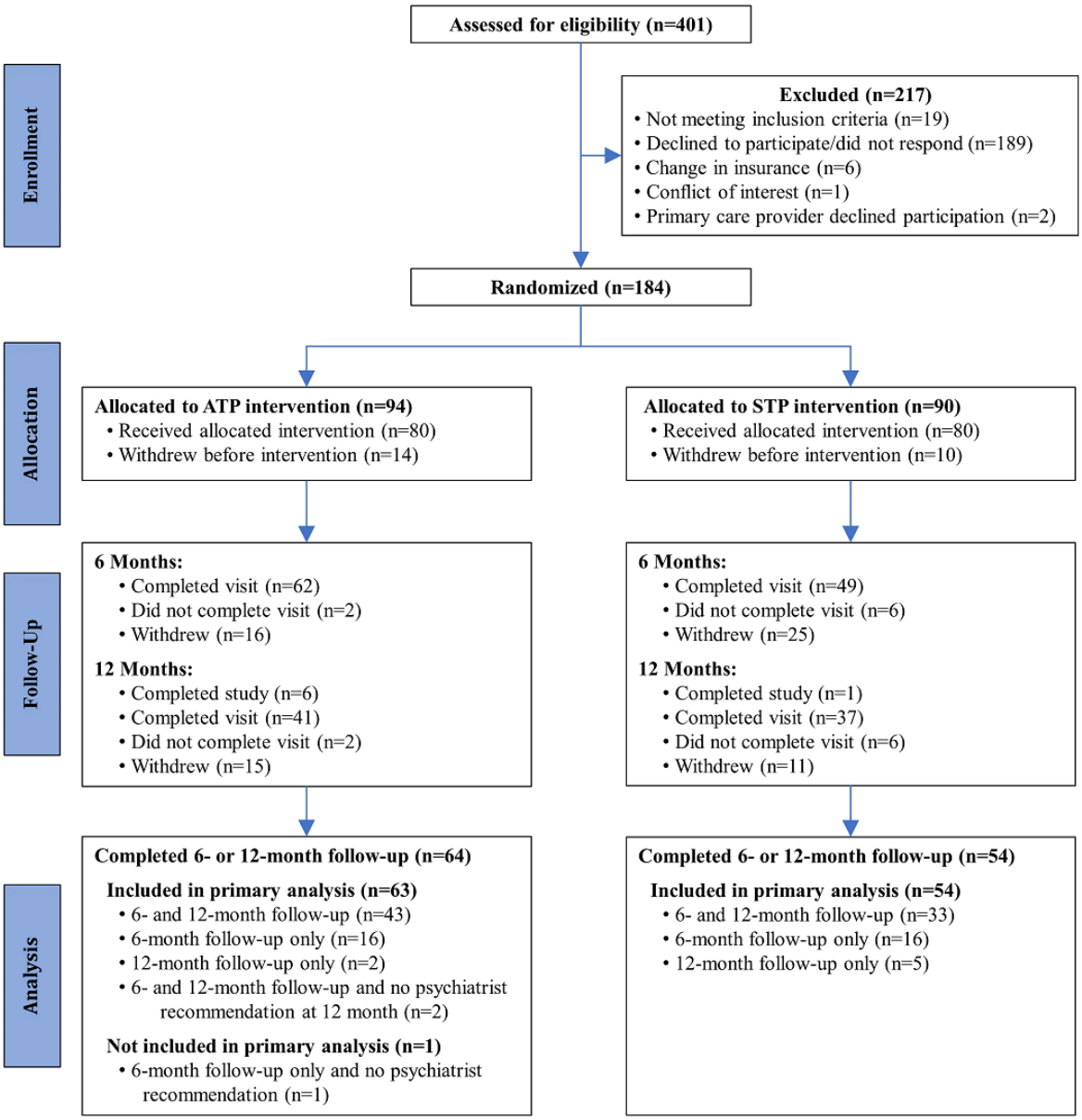


Figure S1 in Multimedia Appendix 1 depicts the flow of patients from screening through the last assessment of the study (24-month follow-up). The dropout rates were higher than originally anticipated. From baseline to the end of the study, 64% (51/80) patients in the ATP arm and 78% (62/80) patients in the STP arm did not complete the study as enrolled (ie, for 12 or 24 months). In the ATP group, 9% (7/80) of the participants had insurance changes, 23% (18/80) declined to continue, 31% (25/80) were lost to follow-up, and 1 did not complete the study because of an administrative scheduling error. In the STP group, 14% (11/80) had insurance changes, 18% (14/80) declined to continue, 43% (34/80) were lost to follow-up, 1 participant was too ill to continue the study, and 2 participants died. There was no association between arm assignment and the reasons for not completing the study (*P*=.33).

Baseline demographic and clinical characteristics by study arm are presented in Table 1. There were no significant arm

XSL•FO
RenderX

differences in any of these characteristics, indicating that the randomization was effective. Of the 160 participants, 137 (85.6%) were White and 30 (18.8%) were Hispanic. Most patients (54 ATP, 54 STP, ie, 67% for each arm) had depressive disorders as their primary diagnosis. Of the 160 patients with baseline data, 117 (73.1%) completed follow-up at 6 months or 12 months and had usable primary outcome data (CGI or GAF) for at least 1 follow-up and were included in the primary analyses (Figure 1). Multimedia Appendix 1, Table S2 compares the demographic and clinical characteristics of the participants who completed the baseline visit and were included (n=117) or excluded (n=43) from the primary analyses. Participants excluded from the primary analyses were more likely to have been patients in the Auburn clinic and had lower CGI scores and higher GAF scores at baseline (Multimedia Appendix 1, Table S2).

XSL•FO
RenderX

Table 1. Baseline demographic and clinical characteristics of participants who completed baseline visits[a].

Characteristics[b]	ATP[c] (n=80)	STP[d] (n=80)	Total (n=160)
Age (years), mean (SD)	53 (14)	52.2 (14.6)	52.6 (14.3)
Number of axis I diagnoses, mean (SD)	2.4 (1)	2.4 (1)	2.4 (1)
Screening PHQ-9[e] score[f,g], mean (SD)	13.9 (6.6)	13.1 (5.9)	13.5 (6.3)
Screening PHQ-9 category[g], n (%)			
0-4, nondepressed	5 (6.3)	8 (10.3)	13 (8.3)
5-9, mild depression	18 (22.8)	16 (20.5)	34 (21.7)
10-14, moderate depression	21 (26.6)	23 (29.5)	44 (28)
≥15, moderately severe to severe depression	35 (44.3)	31 (39.7)	66 (42)
Primary diagnosis, n (%)			
Mood disorder	54 (67.5)	54 (67.5)	108 (67.5)
Anxiety disorder	16 (20)	16 (20)	32 (20)
Substance abuse	2 (2.5)	1 (1.3)	3 (1.9)
Other	8 (10)	9 (11.3)	17 (10.6)
Female, n (%)	58 (72.5)	53 (66.3)	111 (69.4)
Hispanic ethnicity, n (%)	15 (18.8)	15 (18.8)	30 (18.8)
Education, n (%)			
Graduate high school or less	22 (27.5)	18 (22.5)	40 (25)
Some college or 2-year college	32 (40)	40 (50)	72 (45)
College or graduate school	26 (32.5)	22 (27.5)	48 (30)
Marital status[h], n (%)			
Married or living with someone	39 (51.3)	39 (52.7)	78 (52)
Other[i]	37 (48.7)	35 (47.3)	72 (48)
Current psychiatric treatment[j], n (%)	31 (39.7)	34 (42.5)	65 (41.1)
Current psychotropic medication[k], n (%)	64 (83.1)	66 (82.5)	130 (82.8)
Language of the interview, n (%)			
English	71 (88.8)	70 (87.5)	141 (88.1)
Spanish	9 (11.3)	10 (12.5)	19 (11.9)
Study clinic, n (%)			
Auburn	44 (55)	43 (53.8)	87 (54.4)
J Street (Sacramento)	17 (21.3)	19 (23.8)	36 (22.5)
Communicare	19 (23.8)	18 (22.5)	37 (23.1)

[a]Due to rounding, percentages might not sum to 100.

[b]There were no significant differences between the 2 intervention groups for any characteristic.

[c]ATP: asynchronous telepsychiatry.

[d]STP: synchronous telepsychiatry.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]Range 0-27, higher is more depressed.

[g]Data missing=1 in asynchronous telepsychiatry group and 2 in synchronous telepsychiatry.

[h]Data missing=4 in asynchronous telepsychiatry group and 6 in synchronous telepsychiatry.

[i]Includes widowed, divorced or annulled, separated, and never married.

[j]Data missing=2 in asynchronous telepsychiatry group.

[k]Data missing=3 in asynchronous telepsychiatry group.

XSL•FO
RenderX

Primary Outcomes

Table 2 summarizes the mean trajectories and changes from baseline in the 2 intervention arms for the clinician ratings (CGI and GAF) and the results of mixed-effects models for the primary analysis. For both CGI and GAF, ATP did not promote greater improvement than STP at the 6- (ATP vs STP, adjusted difference in follow-up at 6 months vs baseline differences for CGI: 0.2, 95% CI −0.2 to 0.6; *P*=.28; and GAF: −0.6, 95% CI −3.1 to 1.9; *P*=.66) or 12-month follow-up (ATP vs STP, adjusted difference in follow-up at 12 months vs baseline differences for CGI: 0.4, 95% CI −0.04 to 0.8; *P*=.07; and GAF: −0.5, 95% CI −3.3 to 2.2; *P*=.70). However, both the ATP and STP arms improved at 6 and 12 months compared with baseline. Patients in both arms had about 1 point improvement in CGI at 6-month follow-up (estimated difference from baseline −0.7, 95% CI −1.0 to −0.4; *P*<.001 for ATP; and −0.9, 95% CI −1.2 to −0.6; *P*<.001 for STP), and these improvements were maintained at 12 months (estimated difference from baseline −0.8, 95% CI −1.1 to −0.5; *P*<.001 for ATP; and −1.2, 95% CI −1.5 to −0.9; *P*<.001 for STP). The results for GAF were similar, with both groups improving by about 3 points at 6-month (estimated difference from baseline 2.7, 95% CI 1.1-4.4; *P*=.002 for ATP; and 3.3, 95% CI 1.4-5.1; *P*<.001 for STP) and by about 5 points at 12-month follow-up (estimated difference from baseline 4.7, 95% CI 2.8-6.5; *P*<.001 for ATP; and 5.2, 95% CI 3.2-7.2; *P*<.001 for STP).

Table 2. Primary outcomes: clinician ratings at baseline and 6- and 12-month follow-up for the 117 patients included in the primary analysis.

Primary outcomes	Patient, n (%)	CGI[a], mean (SD)	GAF[b], mean (SD)	CGI; estimate, mean (95% CI)[c]	GAF; estimate, mean (95% CI)[c]
ATP[d]					
Mean trajectory					
Baseline	63 (100)	3.9 (0.9)	59.7 (10.8)	__[e]	—
Follow-up at 6 months	61 (97)	3.2 (1)	62.4 (11.9)	—	—
Follow-up at 12 months	45 (71)	3.1 (1.1)	63.7 (13)	—	—
Change from baseline					
6 months versus baseline	61 (97)	−0.7 (1)	2.8 (6.3)	−0.7 (−1.0 to −0.4)	2.7 (1.1 to 4.4)
12 months versus baseline	45 (71)	−0.8 (1.2)	4.4 (8.7)	−0.8 (−1.1 to −0.5)	4.7 (2.8 to 6.5)
STP[f]					
Mean trajectory					
Baseline	54 (100)	4.2 (1)	57.6 (10.2)	—	—
Follow-up at 6 months	49 (91)	3.3 (1)	60.7 (11.0)	—	—
Follow-up at 12 months	38 (70)	3.0 (1)	61.8 (12.2)	—	—
Change from baseline					
6 months versus baseline	49 (91)	−0.9 (1)	2.9 (6.4)	−0.9 (−1.2 to −0.6)	3.3 (1.4 to 5.1)
12 months versus baseline	38 (70)	−1.2 (1)	5.1 (6.3)	−1.2 (−1.5 to −0.9)	5.2 (3.2 to 7.2)
ATP versus STP, difference at baseline	—	—	—	−0.3 (−0.6 to 0.1)	0.9 (−2.1 to 4)
ATP versus STP, difference at follow-up at 6 months	—	—	—	−0.1 (−0.4 to 0.3)	0.4 (−2.8 to 3.5)
ATP versus STP, difference at follow-up at 12 months	—	—	—	0.1 (−0.3 to 0.5)	0.4 (−2.9 to 3.7)
ATP versus STP, difference in follow-up at 6 months versus baseline differences	—	—	—	0.2 (−0.2 to 0.6)	−0.6 (−3.1 to 1.9)
ATP versus STP, difference in follow-up at 12 months versus baseline differences	—	—	—	0.4 (−0.04 to 0.8)	−0.5 (−3.3 to 2.2)

[a]CGI: Clinical Global Impression scale; severity of illness; range 1-7, higher is more severe.

[b]GAF: Global Assessment of Functioning; range 0-100, higher is better functioning.

[c]From mixed-effects linear regression models adjusted for study site, consulting psychiatrist, and language of the interview, as well as clustering due to patient. The model for the Global Assessment of Functioning was further adjusted for clustering by the referring physician.

[d]ATP: asynchronous telepsychiatry.

[e]Not available.

[f]STP: synchronous telepsychiatry.

XSL•FO
RenderX

Secondary Outcomes

Tables 3 and 4 show the descriptive statistics and the results of mixed-effects models for patient self-reported ratings: PHS-12, MHS-12, and PHQ-9, respectively. The pattern of the self-reported ratings was less consistent throughout the follow-up for both ATP and STP arms, with only the mental health score in STP showing statistically significant improvement at 6 months and the PHQ-9 score showing improvement in the ATP group at both 6 and 12 months. However, there were no statistically significant differences in improvement between the intervention arms at any time point for any patient self-reported ratings.

Table 3. Secondary outcomes: patient self-reported 12-Item Short-Form Health Survey (physical and mental) scores at baseline and 6- and 12-month follow-up for the 117 patients included in the primary analysis.

Secondary outcomes	Patient, n (%)	PHS-12[a], mean (SD)	MHS-12[b], mean (SD)	PHS-12; estimate, mean (95% CI)[c]	MHS-12; estimate, mean (95% CI)[c]
ATP[d]					
Mean trajectory					
Baseline	52 (83)	39.6 (11.6)	34.4 (9.6)	—[e]	—
Follow-up at 6 months	51 (81)	39.5 (11.5)	36.7 (9.8)	—	—
Follow-up at 12 months	42 (67)	38.7 (11.5)	38.2 (9.1)	—	—
Change from baseline					
6 months versus baseline	43 (68)	−1.4 (8.8)	2 (11.9)	−1.2 (−3.9 to 11.61.6)	2.5 (−0.7 to 5.7)
12 months versus baseline	33 (52)	0.3 (9.3)	3.7 (12.5)	0.1 (−3.0 to 3.2)	3.6 (−0.003 to 7.1)
STP[f]					
Mean trajectory					
Baseline	45 (83)	43.4 (10.4)	31.7 (8.9)	—	—
Follow-up at 6 months	41 (76)	41.3 (10.5)	36 (11.1)	—	—
Follow-up at 12 months	28 (52)	43.9 (9.4)	34.3 (10.4)	—	—
Change from baseline					
6 months versus baseline	34 (63)	−1.8 (11.4)	5.1 (10.4)	−2.1 (−5 to 0.8)	4.7 (1.4 to 8.1)
12 months versus baseline	24 (44)	−1.1 (8.9)	5 (9.9)	0.001 (−3.3 to 3.3)	3.7 (-0.2 to 7.5)
ATP versus STP, difference at baseline	—	—	—	−9.5 (−32.5 to 13.6)	−2.7 (−24.1 to 18.8)
ATP versus STP, difference at follow-up at 6 months	—	—	—	−8.6 (−31.5 to 14.4)	−4.9 (−26.1 to 16.3)
ATP versus STP, difference at follow-up at 12 months	—	—	—	−9.4 (−32.5 to 13.8)	−2.8 (−24.4 to 18.8)
ATP versus STP, difference in follow-up at 6 months versus baseline differences	—	—	—	0.9 (−3.1 to 4.9)	−2.2 (−6.9 to 2.5)
ATP versus STP, difference in follow-up at 12 months versus baseline differences	—	—	—	0.1 (−4.4 to 4.7)	−0.1 (−5.3 to 5.1)

[a]PHS-12: 12-Item Short-Form Health Survey Physical Health Summary; range 0-100, higher is better physical health.

[b]MHS-12: 12-Item Short-Form Health Survey Mental Health Summary; range 0-100, higher is better mental health.

[c]From mixed-effects linear regression models adjusted for study site, consulting psychiatrist, and language of the interview as well as clustering due to patient and primary care physician.

[d]ATP: asynchronous telepsychiatry.

[e]Not available.

[f]STP: synchronous telepsychiatry.

XSL•FO
RenderX

Table 4. Secondary outcomes: patient self-reported Patient Health Questionnaire-9 scores at baseline and 6- and 12-month follow-up for the 117 patients included in the primary analysis.

Secondary outcomes	Patient, n (%)	PHQ-9[a], mean (SD)	PHQ-9; estimate, mean (95% CI)[b]
ATP[c]			
Mean trajectory			
Baseline	61 (97)	12.4 (7.2)	—[d]
Follow-up at 6 months	57 (90)	9.8 (6.7)	—
Follow-up at 12 months	45 (71)	10 (6)	—
Change from baseline			
6 months versus baseline	55 (87)	−2.3 (4.4)	−2.4 (−3.8 to −0.9)
12 months versus baseline	43 (68)	−2.8 (5.2)	−2.2 (−3.9 to −0.5)
STP[e]			
Mean trajectory			
Baseline	53 (98)	12.6 (6.8)	—
Follow-up at 6 months	40 (74)	10.8 (6.5)	—
Follow-up at 12 months	34 (63)	11.9 (7.1)	—
Change from baseline			
6 months versus baseline	40 (74)	−0.7 (4.8)	−0.9 (−2.5 to 0.8)
12 months versus baseline	33 (61)	−0.5 (6.4)	−0.7 (−2.4 to 1.0)
ATP versus STP, difference at baseline	—	—	1.8 (−9.4 to 13.1)
ATP versus STP, difference at follow-up at 6 months	—	—	0.3 (−10.9 to 11.6)
ATP versus STP, difference at follow-up at 12 months	—	—	0.3 (−11 to 11.6)
ATP versus STP, difference in follow-up at 6 months versus baseline differences	—	—	−1.5 (−3.7 to 0.6)
ATP versus STP, difference in follow-up at 12 months versus baseline differences	—	—	−1.5 (−3.9 to 0.9)

[a]PHQ-9: Patient Health Questionnaire-9; range 0-27, higher is more depressed.

[b]From mixed-effects linear regression models adjusted for study site, consulting psychiatrist, and language of the interview, as well as clustering due to patient and primary care physician.

[c]ATP: asynchronous telepsychiatry.

[d]Not available.

[e]STP: synchronous telepsychiatry.

The results of the secondary analysis (Multimedia Appendix 1, Tables S3-S5) parallel those of the primary analysis, with ATP and STP groups maintaining improvements in both CGI and GAF at 18 and 24 months as compared with baseline and showed no significant interactions between the intervention arm and follow-up times. Sensitivity analyses adjusted for baseline score severity confirmed the results of the primary analyses.

Treatment Adherence, Data Availability, and Unanticipated Events

Adverse or unanticipated events during the trial were reported to the IRB and the DSMB and included 2 patient deaths from unrelated medical complications and 2 patients who threatened self-harm. Both patients who threatened self-harm were urgently contacted by study psychiatrists to make clinical decisions on their follow-up care. The DSMB determined that neither event was study related. A total of 2 participants were randomized to ATP but completed the STP procedures after the baseline. One of these patients, who was urgently seen in person because of suicidal ideation, requested to continue seeing a psychiatrist through STP and was switched from ATP to STP. Another patient was misscheduled from ATP to STP and continued the study in the STP group. During the course of the study, for administrative reasons, the study psychiatrists failed to return notes for 10 completed visits (8 ATP and 2 STP).

Discussion

Principal Findings

This study is the first randomized controlled clinical trial to compare STP with ATP in primary care. It has a number of strengths, including being conducted in primary care practice settings using a collaborative care model (3 sites, including 1 rural site) with a diverse patient sample of English- and

XSL•FO
RenderX

Spanish-speaking patients and a much longer follow-up period than most psychiatric clinical trials. Clinical outcomes were assessed using both clinician and patient self-reported ratings and included both mental health as well as broader measures of health status. A large number of PCPs continued to refer patients to the study for 4 years. The referred patients were typically individuals with mild to moderate anxiety and depression who were mainly treated by PCPs and often did not receive care from a psychiatrist.

At both 12 and 24 months of follow-up, we found that ATP was not superior to STP in improving patient outcomes. However, both ATP and STP patients showed improvements from baseline in 2 separate clinician-rated outcomes at 12-month (of about 1 point for functional impairment on the CGI and 5 points on symptom severity for the GAF) and 24-month follow-up (of about 1 point for CGI and 8 points for GAF). The magnitude of these improvements is similar to those found in recent clinical trials on the effect of nonpharmacological interventions on patient outcomes [45,49,50]. Studies suggest that the minimum clinically meaningful change on the CGI is a 1-point change [49,50], and we have not found similar studies using the CGI outcome for follow-up periods longer than 6 months. A 1-point improvement in our relatively mildly ill population, as we found, is arguably even more clinically significant than in a population that was more severely ill on average at baseline. The findings of an improvement of 8 points on the GAF are similar to long-term therapies in comparable clinical trials [51].

This study was not a noninferiority trial; the sponsor hoped to demonstrate the superiority of ATP. The results did not support the primary hypothesis that ATP promotes more improvement than STP. For clinician ratings, both the ATP and STP arms improved at similar rates throughout the trial, with no significant differences in improvements between the 2 arms. Although not supporting our primary hypothesis, this is still a clinically important result. The standardized implementation of ATP across several primary care settings for a long follow-up period, with improved clinical outcomes at 1 year and 2 years, supports the feasibility of the ATP model of care to treat depression and anxiety using a psychiatric consultation model in patients treated in primary care. This treatment option may be particularly important after the COVID-19 pandemic.

The mental health care system has been significantly affected by the COVID-19 pandemic, with what has been described as a follow-on mental health pandemic [52]. Both the World Health Organization [53] and the Centers for Disease Control [54] have published reports describing greater community levels of depression, anxiety, substance use, domestic violence, sexual abuse and related trauma, and likely suicides. Mental health professionals are required to develop new telepsychiatry protocols and digital systems to help patients who stay at home [55], whereas the number of STP consultations nationwide has dramatically escalated. ATP can provide an innovative solution to treat people in their homes as part of the COVID-19 pandemic response, and the ATP collaborative care model leverages the expertise of psychiatrists to oversee the treatment of larger numbers of patients. In 2020, we continued testing ATP methods to treat patients in their homes and nursing homes. We plan to conduct an ATP homecare trial treating psychiatric disorders in patients who have been severely affected by COVID-19.

The results of this trial have several other implications beyond COVID-19 for broadening access to psychiatric care within underserved populations and across different countries and language groups as well as in reaching new care settings.

First, the results establish that this type of consultation is worth considering as a care option in any collaborative care program. We believe that this large trial of patients treated with ATP for up to 2 years provides evidence that should enable insurers and payers to support payment for ATP consultations. We are already conducting a similar study in skilled nursing facilities, and early engagement and feedback is positive [34]. We see many more opportunities for ATP consultations for assessment and monitoring to occur not only in primary care and remotely in patients' homes but also in pediatric and geriatric psychiatry and correctional environments and for a range of specific psychiatric assessments, such as before bariatric or transplant surgery.

Second, 19.6% (36/184) of the patients enrolled in this trial were Hispanic. It is evident that ATP with a Spanish-speaking interviewer can improve access to psychiatric care for patients who speak only Spanish. This language-matching option provides an important opportunity to increase the availability of mental health care for patients from many language groups within the United States and around the world. We are developing automated language transcription and translation systems to enable this option.

Third, ATP consultations should improve access to psychiatrists. Psychiatrists are in short supply, especially in the primary care environment, and ATP consultations for monitoring and treating patients is an approach to diminishing the impending shortage of these specialists. The 4 treating psychiatrists in this trial did most of their consultations from their usual office environment and were often able to complete their ATP consultations in *downtime* when other patients had canceled. The psychiatrists saw this as a major practical advantage of ATP, as it increased their work efficiency while guiding PCPs more quickly than typical in-person or STP consultations. We are currently undertaking an economic evaluation of trial results. Additional data from this trial will be used to evaluate patient and provider satisfaction, cost-effectiveness, and PCP adherence with the psychiatrist's recommendations and investigate diagnosis-specific clinical outcomes in future publications.

Limitations

There were some limitations to this study. We anticipated a 25% dropout rate for the 2-year study. However, of the 160 patients who completed baseline, 67 (41.9%; 31 ATP and 36 STP) withdrew from the study at either 6 or 12 months, and only 47 (29.4%) patients completed the study as enrolled, despite regular communication from the research team. This is not unusual, as dropout rates in long-term randomized trials for depression in primary care range from 25% to 52% at 1 year [56] and higher for longer studies [57]. Although our dropout rate was higher than initially anticipated, it was comparable with or even less than that of other similar longitudinal trials.

XSL•FO
RenderX

In a study, only 43.9% of participants had data collected at 12 months [54], whereas our study retained 58.1% of participants (93/160) at 12 months. In another study, 51% of participants completed their 12-month checkup, and 19.6% dropped out of the intervention group with less than 4 weeks of participation [55]. Of the 41.9% (67/160) of patients who completed baseline interviews and withdrew from our study in the first year, some patients reported that they dropped out because they felt good and needed no further treatment, and others because they saw no improvement. Others dropped out because they moved, whereas cessation of coverage by an insurer midway through the trial forced a number of patients (n=18) to seek care elsewhere. Patients were recruited from primary care in Northern California, primarily experiencing depression or mood disorders. Although this population is very socially and ethnically diverse, with more than 100 languages spoken in the Sacramento region, the generalizability of our findings to other settings and types of patients is unknown. Due to the nature of the intervention, blinding for either patients or clinicians was not feasible. Finally, relatively high dropout rates in both arms may have skewed follow-up outcomes if there is a relationship between the propensity of a data point to be missing and its values, although it is difficult to predict the direction of the bias.

Conclusions

Although this clinical trial with a 2-year follow-up period does not provide evidence for the superiority of ATP in improving clinical outcomes in comparison with STP, there was a significant improvement in primary outcomes in patients treated with either ATP or STP. Both ATP and STP promise to be important components of collaborative care systems that can increase access to psychiatrists; while ATP, because of its scalability, can improve the efficiency of psychiatric care and help alleviate the shortage of psychiatrists. The bilingual utility of ATP also shows its potential to reach non-English-speaking populations in the United States. Further research could examine the effectiveness of ATP with additional populations, settings, and cost considerations.

Acknowledgments

This work was supported by the Agency for Healthcare Research and Quality through award HS21477 (to PMY). Data management for this project was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through the award UL1 TR001860. The authors thank the members of the Data Safety Monitoring Board, D Dierks, MD; B Durbin-Johnson, PhD; and E Ochoa, MD, and the primary care practitioners who referred their patients to our study and the participants of the study. A Odor, MD, an original investigator on the trial, is deceased.

Authors' Contributions

PMY, MBP, and AMI conceptualized the study. PMY, MBP, and AMI drafted the original draft. AMI designed the analytical plan. AMI, AF, and JRF conducted the analyses. All authors reviewed and approved the final manuscript.

Conflicts of Interest

The authors report no conflicting financial relationships with commercial interests except PMY, who received royalties from the American Psychiatric Association Press. JS is Chief Medical Officer for AccessCare, a provider of telemental health services, and has received royalties from the American Psychiatric Association Press and Springer Press. AMI has received honoraria for reviewing activities from Elsevier. SRC provided consultation services for Advanced Clinical, University of Wisconsin, Madison, and Orbit Health Telepsychiatry; he has also provided speaking services for North American Center for Continuing Medical Education, LLC; Scholastic Expeditions; Arizona Psychiatric Society; and Guidewell Innovation. He has affiliations with University of California, San Francisco; University of California, Davis; and Stanford University School of Medicine.

Multimedia Appendix 1
Additional data and materials.
[DOCX File , 627 KB - jmir_v23i7e24047_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1154 KB - jmir_v23i7e24047_app2.pdf]

References

1. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The empirical evidence for telemedicine interventions in mental disorders. Telemed J E Health 2016 Feb;22(2):87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]
2. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
3. Elliott T, Yopes MC. Direct-to-consumer telemedicine. J Allergy Clin Immunol Pract 2019 Nov;7(8):2546-2552. [doi: 10.1016/j.jaip.2019.06.027] [Medline: 31706486]

XSL•FO
RenderX

4. Shore JH, Schneck CD, Mishkind MC. Telepsychiatry and the coronavirus disease 2019 pandemic-current and future outcomes of the rapid virtualization of psychiatric care. JAMA Psychiatry 2020 Dec 1;77(12):1211-1212. [doi: 10.1001/jamapsychiatry.2020.1643] [Medline: 32391861]
5. Psychiatrists Use of Telepsychiatry During COVID-19 Public Health Emergency: Policy Recommendations. American Psychiatric Association. 2020. URL: https://www.psychiatry.org/psychiatrists/practice/telepsychiatry/blog/apa-resources-on-telepsychiatry-and-covid-19 [accessed 2021-06-22]
6. Monthly Telehealth Regional Tracker. Fair Health. Amazon News. URL: https://s3.amazonaws.com/media2.fairhealth.org/infographic/telehealth/dec-2020-national-telehealth.pdf [accessed 2021-06-22]
7. Ambulatory Care Use and Physician Office Visits. National Center for Health Statistics. 2020. URL: https://www.cdc.gov/nchs/fastats/physician-visits.htm [accessed 2021-06-22]
8. Yellowlees P, Shore J. Telepsychiatry and Health Technologies: a Guide for Mental Health Professionals. Washington, DC: The American Psychiatric Association; 2018.
9. Shore JH, Yellowlees P, Caudill R, Johnston B, Turvey C, Mishkind M, et al. Best practices in videoconferencing-based telemental health april 2018. Telemed J E Health 2018 Nov;24(11):827-832. [doi: 10.1089/tmj.2018.0237] [Medline: 30358514]
10. Brooks E, Turvey C, Augusterfer EF. Provider barriers to telemental health: obstacles overcome, obstacles remaining. Telemed J E Health 2013 Jun;19(6):433-437. [doi: 10.1089/tmj.2013.0068] [Medline: 23590176]
11. Cowan K, McKean A, Gentry M, Hilty D. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
12. Nesbitt T, Yellowlees P, Hogarth M, Hilty D. Quality Through Collaboration: The Future of Rural Health. Washington DC: The National Academies Press; 2005.
13. Collins K, Nicolson P, Bowns I, Walters S. General practitioners' perceptions of store-and-forward teledermatology. J Telemed Telecare 2000;6(1):50-53. [doi: 10.1258/1357633001933844] [Medline: 10824392]
14. Chan S, Li L, Torous J, Gratzer D, Yellowlees PM. Review of use of asynchronous technologies incorporated in mental health care. Curr Psychiatry Rep 2018 Aug 28;20(10):85-11. [doi: 10.1007/s11920-018-0954-3] [Medline: 30155593]
15. Thrall JH. Teleradiology. Part I. History and clinical applications. Radiology 2007 Jun;243(3):613-617. [doi: 10.1148/radiol.2433070350] [Medline: 17517922]
16. High WA, Houston MS, Calobrisi SD, Drage LA, McEvoy MT. Assessment of the accuracy of low-cost store-and-forward teledermatology consultation. J Am Acad Dermatol 2000 May;42(5 Pt 1):776-783. [doi: 10.1067/mjd.2000.104519] [Medline: 10775853]
17. Rashid E, Ishtiaq O, Gilani S, Zafar A. Comparison of store and forward method of teledermatology with face-to-face consultation. J Ayub Med Coll Abbottabad 2003;15(2):34-36. [Medline: 14552246]
18. Rotvold G, Knarvik U, Johansen MA, Fossen K. Telemedicine screening for diabetic retinopathy: staff and patient satisfaction. J Telemed Telecare 2003;9(2):109-113. [doi: 10.1258/135763303321327984] [Medline: 12699582]
19. Hooper GS, Yellowlees P, Marwick TH, Currie PJ, Bidstrup BP. Telehealth and the diagnosis and management of cardiac disease. J Telemed Telecare 2001;7(5):249-256. [doi: 10.1258/1357633011936471] [Medline: 11571078]
20. Mahnke CB, Mulreany MP, Inafuku J, Abbas M, Feingold B, Paolillo JA. Utility of store-and-forward pediatric telecardiology evaluation in distinguishing normal from pathologic pediatric heart sounds. Clin Pediatr (Phila) 2008 Nov;47(9):919-925. [doi: 10.1177/0009922808320596] [Medline: 18626106]
21. Williams S, Henricks WH, Becich MJ, Toscano M, Carter AB. Telepathology for patient care: what am I getting myself into? Adv Anat Pathol 2010 Mar;17(2):130-149. [doi: 10.1097/PAP.0b013e3181cfb788] [Medline: 20179435]
22. Giansanti D, Castrichella L, Giovagnoli MR. Telepathology requires specific training for the technician in the biomedical laboratory. Telemed J E Health 2008 Oct;14(8):801-807. [doi: 10.1089/tmj.2007.0130] [Medline: 18954250]
23. Richardson CG, Slemon A, Gadermann A, McAuliffe C, Thomson K, Daly Z, et al. Use of asynchronous virtual mental health resources for covid-19 pandemic-related stress among the general population in Canada: cross-sectional survey study. J Med Internet Res 2020 Dec 30;22(12):e24868 [FREE Full text] [doi: 10.2196/24868] [Medline: 33315583]
24. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
25. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
26. Melmed A. Chat with a doctor: using asynchronous virtual care access for on-demand physician advice. Iproc 2017 Sep 22;3(1):e18 [FREE Full text] [doi: 10.2196/iproc.8451]
27. Myers K, Cummings JR, Zima B, Oberleitner R, Roth D, Merry SM, et al. Advances in asynchronous telehealth technologies to improve access and quality of mental health care for children and adolescents. J Technol Behav Sci 2018 Apr 6;3(2):87-106. [doi: 10.1007/s41347-018-0055-5]
28. Balasinorwala VP, Shah NB, Chatterjee SD, Kale VP, Matcheswalla YA. Asynchronous telepsychiatry in maharashtra, India: study of feasibility and referral pattern. Indian J Psychol Med 2014 Jul 1;36(3):299-301 [FREE Full text] [doi: 10.4103/0253-7176.135384] [Medline: 25035555]

XSL•FO
RenderX

29. Odor A, Yellowlees P, Hilty D, Parish MB, Nafiz N, Iosif A. PsychVACS: a system for asynchronous telepsychiatry. Telemed J E Health 2011 May;17(4):299-303. [doi: 10.1089/tmj.2010.0159] [Medline: 21457014]
30. Yellowlees PM, Odor A, Parish MB, Iosif A, Haught K, Hilty D. A feasibility study of the use of asynchronous telepsychiatry for psychiatric consultations. Psychiatr Serv 2010 Aug;61(8):838-840. [doi: 10.1176/ps.2010.61.8.838] [Medline: 20675845]
31. Yellowlees P, Burke Parish M, González A, Chan S, Hilty D, Iosif A, et al. Asynchronous telepsychiatry: a component of stepped integrated care. Telemed J E Health 2018 May;24(5):375-378 [FREE Full text] [doi: 10.1089/tmj.2017.0103] [Medline: 29024612]
32. Yellowlees PM, Odor A, Parish MB. Cross-lingual asynchronous telepsychiatry: disruptive innovation? Psychiatr Serv 2012 Sep 1;63(9):945. [doi: 10.1176/appi.ps.2012op945] [Medline: 22949020]
33. Yellowlees PM, Odor A, Parish MB, Iosif A, Haught K, Hilty D. A feasibility study of the use of asynchronous telepsychiatry for psychiatric consultations. Psychiatr Serv 2010 Aug;61(8):838-840. [doi: 10.1176/ps.2010.61.8.838] [Medline: 20675845]
34. Xiong GL, Iosif A, Godwin HT, Khan M, Parish MB, Yellowlees P, et al. A pilot randomized trial of asynchronous and synchronous telepsychiatry in skilled nursing facilities. J Am Med Dir Assoc 2018 May;19(5):461-462. [doi: 10.1016/j.jamda.2018.02.007] [Medline: 29704928]
35. Butler TN, Yellowlees P. Cost analysis of store-and-forward telepsychiatry as a consultation model for primary care. Telemed J E Health 2012 Jan;18(1):74-77. [doi: 10.1089/tmj.2011.0086] [Medline: 22085113]
36. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psych Annals 2002 Sep;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
37. First M. Structured clinical interview for the DSM (SCID). Clin Psychol 2014:1-6.
38. Parish MB, Gonzalez A, Hilty D, Chan S, Xiong G, Scher L, et al. Asynchronous telepsychiatry interviewer training recommendations: a model for interdisciplinary, integrated behavioral health care. Telemed J E Health 2021 Jan 12:- epub ahead of print. [doi: 10.1089/tmj.2020.0076] [Medline: 33434453]
39. Guy W. Clinical Global Impressions: ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.
40. Hall RC. Global assessment of functioning. Psychosomatics 1995 May;36(3):267-275. [doi: 10.1016/s0033-3182(95)71666-8]
41. Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M, et al. The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. J Eval Clin Pract 2008 Dec;14(6):979-983. [doi: 10.1111/j.1365-2753.2007.00921.x] [Medline: 18462279]
42. Kørner A, Lauritzen L, Abelskov K, Gulmann NC, Brodersen A, Wedervang-Jensen T, et al. Rating scales for depression in the elderly: external and internal validity. J Clin Psychiatry 2007 Mar 15;68(3):384-389. [doi: 10.4088/jcp.v68n0305] [Medline: 17388707]
43. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries. J Clin Epidemio 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7]
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
45. Dunner DL, Aaronson ST, Sackeim HA, Janicak PG, Carpenter LL, Boyadjis T, et al. A multisite, naturalistic, observational study of transcranial magnetic stimulation for patients with pharmacoresistant major depressive disorder. J Clin Psychiatry 2014 Sep 16;75(12):1394-1401. [doi: 10.4088/jcp.13m08977]
46. Müller-Nordhorn J, Roll S, Willich SN. Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. Heart 2004 May;90(5):523-527 [FREE Full text] [doi: 10.1136/hrt.2003.013995] [Medline: 15084550]
47. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982 Dec;38(4):963-974. [Medline: 7168798]
48. SAS/STAT 9.4. 2013. URL: https://www.sas.com [accessed 2021-06-22]
49. Roux P, Brunet-Gouet E, Ehrminger M, Aouizerate B, Aubin V, Azorin JM, FondaMental Advanced Centers of Expertise in Bipolar Disorders (FACE-BD) Collaborators, et al. Minimum clinically important differences for the functioning assessment short test and a battery of neuropsychological tests in bipolar disorders: results from the FACE-BD cohort. Epidemiol Psychiatr Sci 2020 Jul 20;29:e144 [FREE Full text] [doi: 10.1017/S2045796020000566] [Medline: 32684190]
50. Mischoulon D, Hylek L, Yeung AS, Clain AJ, Baer L, Cusin C, et al. Randomized, proof-of-concept trial of low dose naltrexone for patients with breakthrough symptoms of major depressive disorder on antidepressants. J Affect Disord 2017 Jan 15;208:6-14. [doi: 10.1016/j.jad.2016.08.029] [Medline: 27736689]
51. Lorentzen S, Ruud T, Fjeldstad A, Høglend P. Comparison of short- and long-term dynamic group psychotherapy: randomised clinical trial. Br J Psychiatry 2013 Sep 2;203(3):280-287. [doi: 10.1192/bjp.bp.112.113688] [Medline: 24029539]
52. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. The Lancet Psychiatry 2020 Mar;7(3):228-229. [doi: 10.1016/s2215-0366(20)30046-8]
53. Mental Health and Covid-19. World Health Organization. 2020. URL: https://www.who.int/teams/mental-health-and-substance-use/covid-19 [accessed 2021-06-21]

XSL•FO
RenderX

54. Stephenson J. CDS report reveals 'considerably elevated' mental health toll from covid-19 stresses. JAMA Health Forum 2020 Aug 25;1(8):e201078. [doi: 10.1001/jamahealthforum.2020.1078]
55. Yellowlees P, Nakagawa K, Pakyurek M, Hanson A, Elder J, Kales HC. Rapid conversion of an outpatient psychiatric clinic to a 100% virtual telepsychiatry clinic in response to covid-19. Psychiatr Serv 2020 Jul 1;71(7):749-752. [doi: 10.1176/appi.ps.202000230] [Medline: 32460683]
56. Wilkinson P, Izmeth Z. Continuation and maintenance treatments for depression in older people. Cochrane Database Syst Rev 2016 Sep 9;9:CD006727 [FREE Full text] [doi: 10.1002/14651858.CD006727.pub3] [Medline: 27609183]
57. Schulberg HC, Coulehan JL, Block MR, Scott CP, Imber SD, Perel JM. Strategies for evaluating treatments for major depression in primary care patients. Gen Hosp Psychiatry 1991 Jan;13(1):9-18. [doi: 10.1016/0163-8343(91)90004-g] [Medline: 1993523]

Abbreviations

ATP: asynchronous telepsychiatry
CGI: Clinical Global Impression
DSMB: data and safety monitoring board
EMR: electronic medical records
GAF: Global Assessment of Functioning
IRB: institutional review board
MHS-12: 12-Item Short-Form Health Survey Mental Health Summary
PCP: primary care physician
PHQ-9: Patient Health Questionnaire-9
PHS-12: 12-Item Short-Form Health Survey Physical Health Summary
SCID: Structured Clinical Interview for the Diagnostic Statistical Manual
STP: synchronous telepsychiatry
UCD: University of California Davis

Edited by G Eysenbach; submitted 03.09.20; peer-reviewed by R Bailey, P Chand, MW Haun, S Jyothi, A Khoja, A Serrano-Blanco; comments to author 19.11.20; revised version received 26.03.21; accepted 16.05.21; published 20.07.21.

Please cite as:
Yellowlees PM, Parish MB, Gonzalez AD, Chan SR, Hilty DM, Yoo BK, Leigh JP, McCarron RM, Scher LM, Sciolla AF, Shore J, Xiong G, Soltero KM, Fisher A, Fine JR, Bannister J, Iosif AM
Clinical Outcomes of Asynchronous Versus Synchronous Telepsychiatry in Primary Care: Randomized Controlled Trial
J Med Internet Res 2021;23(7):e24047
URL: https://www.jmir.org/2021/7/e24047
doi:10.2196/24047
PMID:33993104

©Peter M Yellowlees, Michelle Burke Parish, Alvaro D Gonzalez, Steven R Chan, Donald M Hilty, Byung-Kwang Yoo, J Paul Leigh, Robert M McCarron, Lorin M Scher, Andres F Sciolla, Jay Shore, Glen Xiong, Katherine M Soltero, Alice Fisher, Jeffrey R Fine, Jennifer Bannister, Ana-Maria Iosif. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using a Real-Time Locating System to Evaluate the Impact of Telemedicine in an Emergency Department During COVID-19: Observational Study

Birju Patel[1], MPH, MD; Stacie Vilendrer[2], MS, MBA, MD; Samantha M R Kling[2], RD, PhD; Ian Brown[3], MS, MD; Ryan Ribeira[3], MPH, MD; Matthew Eisenberg[1], MD; Christopher Sharp[1], MD

[1]Department of Medicine, Stanford University School of Medicine, Palo Alto, CA, United States

[2]Evaluation Sciences Unit, Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States

[3]Department of Emergency Medicine, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Stacie Vilendrer, MS, MBA, MD
Evaluation Sciences Unit, Division of Primary Care and Population Health
Department of Medicine
Stanford University School of Medicine
1265 Welch Road x216
Stanford, CA, 94305
United States
Phone: 1 650 723 1146
Email: staciev@stanford.edu

Abstract

Background: Telemedicine has been deployed by health care systems in response to the COVID-19 pandemic to enable health care workers to provide remote care for both outpatients and inpatients. Although it is reasonable to suspect telemedicine visits limit unnecessary personal contact and thus decrease the risk of infection transmission, the impact of the use of such technology on clinician workflows in the emergency department is unknown.

Objective: This study aimed to use a real-time locating system (RTLS) to evaluate the impact of a new telemedicine platform, which permitted clinicians located outside patient rooms to interact with patients who were under isolation precautions in the emergency department, on in-person interaction between health care workers and patients.

Methods: A pre-post analysis was conducted using a badge-based RTLS platform to collect movement data including entrances and duration of stay within patient rooms of the emergency department for nursing and physician staff. Movement data was captured between March 2, 2020, the date of the first patient screened for COVID-19 in the emergency department, and April 20, 2020. A new telemedicine platform was deployed on March 29, 2020. The number of entrances and duration of in-person interactions per patient encounter, adjusted for patient length of stay, were obtained for pre- and postimplementation phases and compared with *t* tests to determine statistical significance.

Results: There were 15,741 RTLS events linked to 2662 encounters for patients screened for COVID-19. There was no significant change in the number of in-person interactions between the pre- and postimplementation phases for both nurses (5.7 vs 7.0 entrances per patient, P=.07) and physicians (1.3 vs 1.5 entrances per patient, P=.12). Total duration of in-person interactions did not change (56.4 vs 55.2 minutes per patient, P=.74) despite significant increases in telemedicine videoconference frequency (0.6 vs 1.3 videoconferences per patient, P<.001 for change in daily average) and duration (4.3 vs 12.3 minutes per patient, P<.001 for change in daily average).

Conclusions: Telemedicine was rapidly adopted with the intent of minimizing pathogen exposure to health care workers during the COVID-19 pandemic, yet RTLS movement data did not reveal significant changes for in-person interactions between staff and patients under investigation for COVID-19 infection. Additional research is needed to better understand how telemedicine technology may be better incorporated into emergency departments to improve workflows for frontline health care clinicians.

(J Med Internet Res 2021;23(7):e29240) doi:10.2196/29240

XSL•FO
RenderX

KEYWORDS

real-time locating system; wearable electronic devices; telemedicine; COVID-19; delivery of health care; real time; wearable; impact; telehealth; emergency department; observational; transmission; risk; infectious disease

Introduction

The COVID-19 pandemic acutely raised concerns about the risks of nosocomial transmission of infection within the walls of health care systems. Several health care systems experienced outbreaks [1-3], sometimes with far-reaching consequences [4]. Health care workers can become ill from nosocomial transmission [5], potentially due to patient exposure [6] and physical proximity, which cripples the ability of a health system to continue functioning at needed capacity. Personal protective equipment (PPE) has been instrumental to reduce the risk of transmission of respiratory viral illness [7,8], but given initial concerns about supply, other tools to limit exposure have been evaluated.

Telemedicine was one technology deployed by health care systems with the goal of protecting patients and staff during the COVID-19 pandemic, enabling clinicians to provide care remotely to both patients who are at home for outpatient evaluation and also on premises for acute care [9-11]. While it has been hypothesized that telemedicine would limit unnecessary in-person contact and reduce the need for PPE in health care facilities [12], the clinical and practical impact of this technology has not been validated for this purpose. For example, logs of video telecommunications can report system utilization but do not necessarily relate to risk of exposure and transmission of SARS-CoV-2 given variations in physical workflows among clinicians. Moreover, determining the impact of telemedicine on infection spread to employees is limited, as obtaining occupational health data presents legitimate privacy concerns. Direct observation of staff members after telemedicine implementation also poses safety risks during a pandemic and excludes the possibility for a baseline comparison. Yet validating the hypothesis that on-premises telemedicine reduces pathogen exposure is needed, particularly given widespread investments in these technologies for this purpose [10].

A real-time locating system (RTLS) uses a combination of sensors installed in defined locations and locators carried by staff or installed on equipment to track movement. These data may then be visible on a live monitor or recorded on a database log for further analysis, where it may provide operational or research value. For example, an RTLS can be used to measure and improve health care delivery [13] by helping understand patterns of staff movement [14-18] and monitoring operational efficiency [19,20]. The quantitative information from an RTLS can be analyzed to predict patterns in clinical workflows [21]. Recently, RTLSs have been used as a method for contact tracing during COVID-19 [22,23]. However, each RTLS deployment may have unique limitations based on its practice setting, such as partial functionality or local resistance to adoption (eg, limited staff participation with wearing tracking badges), that make its utility less clear as a broader monitoring tool [24,25].

We recognized an opportunity to evaluate whether an RTLS can quantify the impact of the rapid deployment of telemedicine on health care clinician workflows, measured by changes in movement through patient rooms. We aimed to understand (1) the feasibility of linking multimodal data, including RTLS data, to develop measures that summarize relevant physical workflows in a complex environment, and (2) describe the changes in in-person interactions between staff and patients observed after deploying a telemedicine platform in an emergency department early in the COVID-19 pandemic.

Methods

Implementing an RTLS in the Hospital

Stanford Health Care (SHC) launched an RTLS platform (Midmark) in conjunction with the opening of a new hospital building in late 2019. Infrared and radiofrequency sensors were installed in every patient care room, and staff were given RTLS badges to wear alongside their name badge [14,26]. The RTLS system required line of sight between the room sensor and staff member's badge to trigger. The badge emitted a ping every 1 to 3 seconds to convey that the staff member was still present in the room, and the system only logged events lasting longer than 5 seconds. The installation team optimized sensitivity settings of the sensors based on the local geometric configuration and construction materials. Thus, if staff walked forward into a patient room wearing the locator badge clipped appropriately to their uniform, the system captured entrance and duration of time spent in the room. However, the in-room sensors did not trigger when staff members were walking down the corridor. Data were not validated by comparing logged events to visual observation by the institution or vendor except in a limited fashion during the installation process; however, prior work has found very high correlation between RTLS events and direct observation, especially when monitoring was focused on patient rooms [18]. While both nurses and physicians were encouraged to wear the badges, the technological features primarily benefitted nurses in their usual work routines: device alarms were automatically silenced by the RTLS upon entering a room, and security assistance could be summoned discreetly via an unlabeled button on the badge. Physicians who chose to wear badges could also take advantage of this security feature but were not required to participate.

Implementing Telemedicine in the Emergency Department

At the start of the COVID-19 pandemic, SHC formed an operational committee to guide its telemedicine strategy. When the first patient was tested for COVID-19 on March 2, 2020, clinicians could activate a previously deployed communication technology (Cisco Jabber) that was primarily used for messaging between clinicians. While the technology did offer a videoconference feature, it required a staff member within the room to accept the call to initiate two-way communication. Given this limitation, a new platform composed of a consumer-grade tablet (Apple iPad) mounted on stands with wheels and a secure videoconferencing service (Zoom) was

XSL•FO
RenderX

deployed in all patient care rooms throughout the emergency department [11]. To initiate a video call, clinicians activated the telemedicine system using centrally located tablets located outside of patient rooms and were automatically connected to the device at the patients' bedside. This telemedicine system was deployed and tested by technology support staff in the adult emergency room starting March 25, 2020, and was released for use by clinicians on March 29, 2020, when clinical champions were present to demonstrate how to use the system and technology support staff were available for troubleshooting. In addition, digital resources (including video tutorials) with instructions on use were provided.

Before entering the emergency department, patients were screened for any symptoms of influenza-like illness outside of the entrance. They were then triaged into one of three pathways: low-acuity patients meeting the screening criteria for COVID-19 were sent for testing in the garage outside of the emergency department (where the RTLS was not deployed), high-acuity patients meeting the screening criteria were sent to a designated waiting room isolated from the remaining emergency department, and patients not meeting the screening criteria were sent to the primary waiting room. Once inside the emergency department, every patient who was screened for COVID-19 was placed on appropriate isolation precautions and underwent laboratory testing for SARS-CoV-2. Clinicians were able to identify patients who were being screened for COVID-19 by isolation signage on the room entrance, an indicator on the electronic locator board, and from within the electronic health record. Clinicians were free to choose whether they would interact with patients with isolation precautions in person or via telemedicine. Nonetheless, normal standards of care were expected, including physical examination of patients during the encounter.

Developing an Analytics Pipeline and Data Analysis

We identified multiple streams of data necessary to quantify the impact of telemedicine on health care worker movement. First, we extracted the RTLS event logs for movement of nursing and physician staff in each patient care room of the emergency department. RTLS data from staff in other clinical roles, such as radiology technicians and phlebotomists, were not included in this study as these roles were not the target users of the telemedicine technology. In addition, all staff members included in the study had dedicated clinical roles in the emergency department. We also obtained utilization logs from the telemedicine platforms. Finally, we queried our electronic health record system for the isolation status of all emergency department encounters that would indicate a patient needed screening for COVID-19, along with room information, time of rooming, and time of disposition decision. Patients who were never placed in isolation for COVID-19 testing during their evaluation were excluded from the analysis, as these patients were not the target of telemedicine usage.

The RTLS and encounter data were then processed and unified to provide measurements of the primary outcomes—the number of staff entrances per qualifying emergency department encounter and the total duration spent in the patient rooms over each encounter. First, when separate, successive RTLS events linking the same staff member and room number occurred within 30 seconds of one another, these events were combined into a single event. This assumption accounts for the possibility that the line of sight between the staff member's badge and room sensor was briefly interrupted by a shift in the location of people or equipment inside the room. We conducted sensitivity analyses using thresholds of 1 second and 10 seconds and observed no significant difference in findings. The RTLS staff movement data was then merged with data on patient encounters using room names and timestamps extracted from the electronic health record system. Since patients with longer encounters in the emergency department are naturally more likely to have more interactions with health care clinicians, measures for staff entrances and duration per encounter were adjusted by encounter length of stay (time between initial rooming and disposition time). For example, to adjust the number of entrances for each encounter, the raw number of staff entrances was divided by that encounter's length of stay and then multiplied by the average encounter length of stay observed over the entire study period. This method of adjustment allowed for data to be analyzed at different levels of granularity, including individual encounters, daily summaries, and over the study phase. In addition, as uptake of RTLS badges was different among physicians and nurses, we generally present statistics stratified by clinical role and use rates instead of absolute counts to make interpretation more consistent.

To describe telemedicine utilization, videoconference logs from both the Cisco and Zoom systems were summarized into daily activity reports for the entire emergency department, as the technology did not track which telemedicine sessions were linked to specific patient encounters. These daily summaries were divided by the daily emergency department census of patients screened for COVID-19 (derived from encounter data) to gather an approximate number of video conferences per patient encounter.

The final data set was then separated into the pre- and postimplementation phases to understand the change over time. To allow for a similar length of observation to the preimplementation phase from March 2 to March 28, data were collected for the postimplementation phase from March 29 to April 20, 2020. For each outcome, mean (SD) values for encounter-level measures were calculated for each phase, along with *t* tests to evaluate for statistically significant differences between the two phases using an alpha of .05.

Staff and patient data were deidentified prior to being available to the analytics team. This project was not deemed human subjects research by our Institutional Review Board (protocol #55927).

Results

Staff and Patients

During the evaluation period, SHC cared for 6951 patients in its adult emergency department, of which 2662 (38.3%) were evaluated for COVID-19, with an average length of stay of 251 minutes. RTLS badges were worn by 245 unique staff members including 40 out of 99 (40.4%) attending physicians, 8 out of

XSL•FO
RenderX

62 (12.9%) resident physicians, and 197 (100.0%) nurses scheduled for service. Rates of badge use were similar between pre- and postimplementation phases (Table 1). We linked 15,741 RTLS events (a staff member entering a patient's room) to the 2662 encounters of patients undergoing screening for COVID-19.

Table 1. Characteristics of patients, staff, and real-time locating system (RTLS) events in the emergency department stratified by study periods. Staff data report the number of staff wearing RTLS badges compared to those scheduled for at least one shift during that study phase.

Variable	Preimplementation phase	Postimplementation phase
Patients in the emergency department		
All patients evaluated, n	4571	2380
Patients tested for COVID-19, n (%)	1502 (32.9)	1059 (44.4)
Length of stay for patients tested for COVID-19 (minutes), mean (SD)	190 (209)	339 (299)
Proportion of staff scheduled who wore an RTLS badge, n/N (%)		
Total staff	220/328 (67.1)	212/315 (67.3)
Nurses	182/182 (100.0)	174/174 (100.0)
Attending physicians	32/96 (33.3)	33/91 (36.3)
Resident physicians	6/50 (12.0)	5/50 (10.0)
RTLS		
RTLS events for COVID-19–tested patients	5855	9886

Uptake of Telemedicine

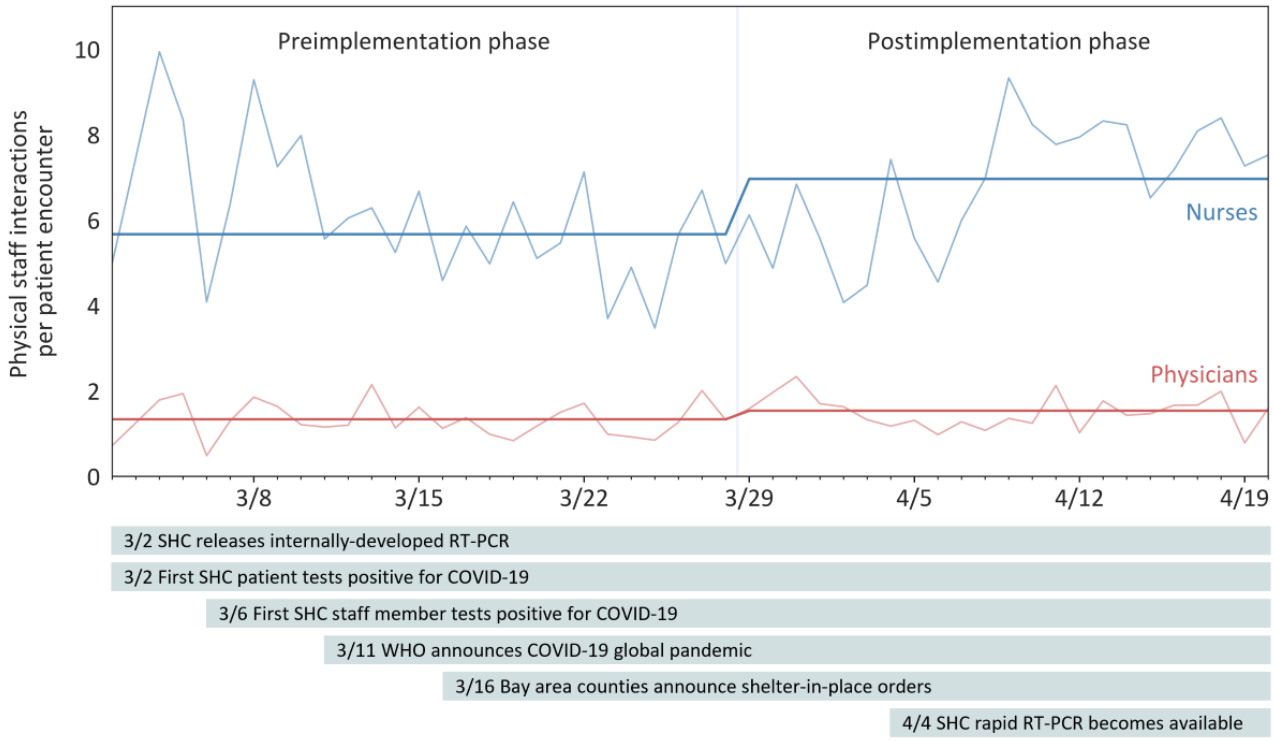
Staff engaged in 892 videoconferences with patients screened for COVID-19 (0.6 videoconferences per patient) during the preimplementation phase using the existing Cisco platform and 1482 videoconferences (1.3 videoconferences per patient) using either the Cisco or Zoom platform during the postimplementation phase ($P<.001$ for change in daily average of videoconferences per patient). The video functionality of the pre-existing communication platform gained use just prior to the deployment of the newer telemedicine platform, with the majority of these videoconferences (796/892, 89.2%) in the preimplementation phase occurring between March 15 and March 29. During the postimplementation phase, the newer Zoom videoconference platform was the dominant technology (906/1482 videoconferences, 61.1%). Usage of telemedicine was highest immediately after deployment (2.0 videoconferences per patient) and decreased in the subsequent weeks (1.4 videoconferences per patient in the second week after deployment and 0.9 videoconferences per patient in the third week; $P=.002$ for the difference between the first and third weeks of the postimplementation phase).

Frequency of Patient Room Entrances

Nurses had an average of 7.3 (SD 6.9) (after adjustment for patient length of stay: mean 5.7, SD 4.1) entrances per patient encounter in the preimplementation phase. After implementation, nurses had an average of 10.8 (SD 10.2) (after adjustment: mean 7.0, SD 5.4) entrances per patient encounter, which was not a statistically significant difference between the phases ($P=.07$, Figure 1). Similarly, physicians had an average of 1.5 (SD 1.1) (after adjustment: mean 1.3, SD 1.0) entrances per patient encounter in the preimplementation phase and 2.2 (SD 2.0) (after adjustment: mean 1.5, SD 1.5) entrances per patient encounter in the postimplementation phase, which was not a statistically significant difference ($P=.12$). Further restricting the analysis to resident physicians also showed a similar pattern. In the preimplementation phase, resident physicians had an average of 1.6 (SD 1.4) (after adjustment: mean 1.1, SD 0.9) entrances per patient encounter, and 2.1 (SD 1.7) (after adjustment: mean 1.4, SD 1.0) entrances per patient encounter, which was not a statistically significant difference between the phases ($P=.22$).

XSL•FO
RenderX

Figure 1. In-person staff interactions with patients under investigation for COVID-19. Entrances into rooms with patients under investigation by nurses (daily census-weighted average over phase in dark blue, daily averages in light blue) had a nonstatistically significant increase between the pre- and postimplementation phases. Physicians (daily census-weighted average over phase in dark red, daily averages in light red) physically entered patient rooms much less often than nurses, and changes in physician entrances into patient rooms over phases were more subtle in absolute counts. A timeline of relevant public health events is provided below [27-29]. SHC: Stanford Health Care, RT-PCR: reverse transcription–polymerase chain reaction, WHO: World Health Organization.

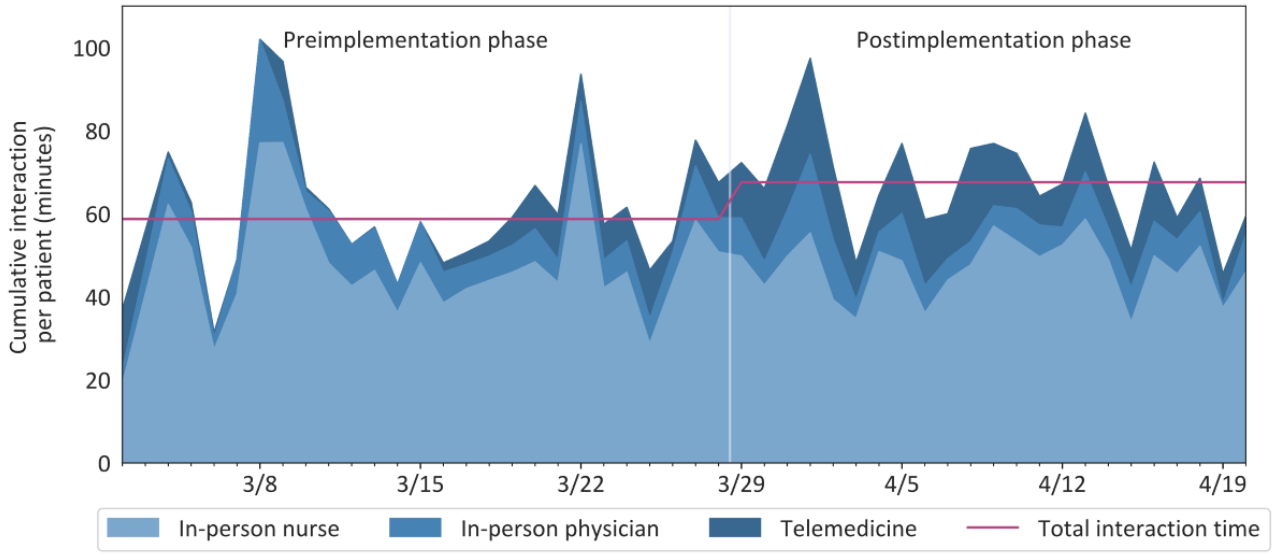


Time Spent With Patient

The combination of videoconferencing and RTLS data created a profile of the staff's mode of interaction with patients. During the preimplementation phase, virtual interactions lasted an average of 4.3 minutes per encounter and accounted for 7.3% of total time staff spent with patients either in the room or via telemedicine (Figure 2). During the postimplementation phase, virtual interactions increased to 12.3 minutes per encounter and accounted for 18.2% of total time staff spent with patients ($P<.01$ for change in daily average). Simultaneously, in-person contact duration remained stable at 78.3 (SD 31.1) (after adjustment for patient length of stay: mean 56.4, SD 17.4) minutes per patient encounter during the preimplementation phase and 78.7 (SD 13.9) (after adjustment: mean 55.2, SD 8.9) minutes per patient encounter during the postimplementation phase, which was not a statistically significant difference ($P=.74$ for change in daily average). In particular, nurses spent 67.1 (SD 25.3) (after adjustment: 48.0, SD 14.3) minutes in person per patient encounter in the preimplementation phase and 67.3 (SD 11.7) (after adjustment: mean 47.2, SD 7.0) minutes in person per patient encounter in the postimplementation phase ($P=.80$), while physicians spent 11.3 (SD 7.8) (after adjustment: mean 8.5, SD 4.2) minutes in person per patient encounter in the preimplementation phase and 11.4 (SD 5.6) (after adjustment: mean 8.0, SD 3.8) minutes in person per patient encounter in the postimplementation phase ($P=.67$). Altogether, combined in-person and virtual interaction time in the postimplementation phase (mean 91.5, SD 14.4 minutes per patient encounter; after adjustment: mean 67.9, SD 11.9) was not statistically different from the preimplementation phase (mean 82.9, SD 30.1 minutes per patient encounter; after adjustment: mean 61.0, SD 17.0; $P=.11$ for change in daily average).

XSL•FO
RenderX

Figure 2. Cumulative staff-patient interaction for patients under investigation for COVID-19. In-person exposure to patients under investigation was mostly borne by nurses (daily average shaded light blue) but did not change significantly over the phases. A similar trend was seen for in-person contact time for physicians (daily average shaded blue). Virtual contact (daily average shaded dark blue) with patients increased during the postimplementation phase, demonstrating adoption of the new telemedicine technology platform beyond the previously available platform. However, there was not a statistically significant difference in total interaction time with patients between the pre- and postimplementation phases (average for entire phase shown as a purple line).



Discussion

Principal Findings

The combination of multimodal data can capture meaningful workflow components of staff movement in an emergency department during the implementation of a new software technology, even in the course of a pandemic. RTLS data have been previously utilized to evaluate clinical workflows using process mining [30] and simulation [31], and this evaluation adds an analytical approach to understand the extent that physical workflows changed after the implementation of a telemedicine technology.

With the addition of a new telemedicine platform, we expected that virtual interactions would replace a portion of in-person contact with patients and thus theoretically reduce pathogen exposure and PPE use. For example, a reduction in the number of entrances into patient care rooms may result in a reduction in PPE use. However, after cross-validating telemedicine adoption data with RTLS data, we found no significant difference in either entrances or duration of in-person physical interaction across multiple clinical roles, including nurses, attending physicians, and resident physicians, suggesting exposure risk was not reduced despite an increased use of videoconferencing. We also found that the cumulative time clinicians spent interacting with patients either in person or virtually did not change significantly in the context of a pandemic and new workflows.

Not unexpectedly and consistent with prior findings [14,32], we observed that nurses, compared to physicians, had the majority of interactions (81% of entrances and 85% of the time) with patients under evaluation for COVID-19. Thus, telemedicine technologies may hold greater potential to impact nursing workflows to decrease transmission risk in a pandemic compared to other health care clinicians; however, additional research is needed to understand what elements of nursing work can be done virtually while maintaining high-quality patient care [33]. For example, while some essential nursing tasks like medication administration need to be done in person, some aspects of communication around those actions could be made virtual. For physicians, several components of their workflows can be completed reasonably using telemedicine [34], but our data suggest that there remain important elements of clinical evaluation, communication, and treatment that physicians prioritize completing in person. Thus, telemedicine may interact with clinical workflows in more complex ways that do not manifest in our presented metrics.

In addition, other concurrent changes in technology and care standards may offer insights into why patient room entrances and time spent in person with the patient did not differ in the pre- and postimplementation of telemedicine technology. First, the development of a rapid SARS-CoV-2 RT-PCR (reverse transcription–polymerase chain reaction) assay early in the postimplementation phase may have allowed physicians to delay in-person evaluations of patients until they received test results, permitting clinicians to modify their routines depending on perceived risk after a negative COVID-19 result. Second, as emergency department staff became more comfortable with isolation protocols, they may have felt less fear of contracting the disease, making them more comfortable spending time in person with patients even when a given clinical activity, such as decision-making or counseling, could have been done via telemedicine. Lastly, during the early period of the pandemic when PPE resources were scarce, nurses may have also consolidated tasks to minimize entrances into isolation rooms. These behavioral and workflow changes may have reverted to baseline levels as the PPE inventory stabilized. However, there was a gap in the timeline between these events both before and after the telemedicine deployment, and telemedicine adoption was greatest immediately after deployment. Thus, the likelihood

XSL•FO
RenderX

of observing a significant change due to the technology was high if that signal truly existed. The lack of significant differences in room entrances and total time spent with patients between the pre- and postimplementation periods may simply be a reflection of this particular technology deployment at a specific phase of the pandemic response, characterized by simultaneous and often transient complexities. Most importantly, an RTLS-based analysis can serve as a useful template for future evaluations of health technologies during emergencies, particularly if a contemporaneous comparator can be found.

Limitations

This evaluation has certain limitations. The RTLS data and telemedicine data were unable to be merged by patient room, limiting our understanding of how staff used telemedicine and how workflow was affected for specific patient groups. Understanding the differing use of telemedicine by nurses and physicians (which was not available in our study) may be a fruitful area for future research, particularly if such data can be matched to individual staff movement data. Finally, as both attending and resident physicians did not universally wear RTLS badges, the physicians included in our analysis may not be representative of others who chose not to wear an RTLS badge. A clear strength of the evaluation, however, is that movement data was available for all nursing staff within a single emergency department and was able to be linked to individual patient encounters.

Conclusion

Movement data captured by an RTLS is a rich technological resource that can assist in monitoring the impact of changes in workflow even in a rapidly changing clinical environment. We were able to leverage an RTLS to quantify staff movement in our emergency department as we deployed a new telemedicine platform during the COVID-19 pandemic. While telemedicine did see a significant increase in adoption in the emergency setting, it did not ultimately influence physician and nursing movement in and out of the rooms of patients who were under investigation for COVID-19. These findings underscore the need for additional formal evaluations to determine whether informatics interventions (and the significant resources directed to their deployment) are having their intended impact on health care worker safety. Such operationally relevant analyses can be enabled by unifying real-world data from multimodal platforms.

Acknowledgments

The authors would like to thank Rajiv Ramdeo, Ann-Marie Yap, Kevin Yuen, Derek Nielsen, Sam Shen, Gretchen Brown, and Steven Asch for their contributions to this project.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. BP and SV were funded through the Clinical Informatics Fellowship and the Stanford-Intermountain Population Health Fellowship, respectively.

Authors' Contributions

All authors made substantial contributions to conception and design as well as acquisition of data or analysis and interpretation of data; were involved in drafting the manuscript or revising it critically for important intellectual content; gave final approval of the version published; and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Rickman HM, Rampling T, Shaw K, Martinez-Garcia G, Hail L, Coen P, et al. Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. Clin Infect Dis 2021 Feb 16;72(4):690-693 [FREE Full text] [doi: 10.1093/cid/ciaa816] [Medline: 32562422]
2. Asad H, Johnston C, Blyth I, Holborow A, Bone A, Porter L, et al. Health Care Workers and Patients as Trojan Horses: a COVID19 ward outbreak. Infection Prevention in Practice 2020 Sep;2(3):100073. [doi: 10.1016/j.infpip.2020.100073]
3. Taylor J, Rangaiah J, Narasimhan S, Clark J, Alexander Z, Manuel R, et al. Nosocomial COVID-19: experience from a large acute NHS Trust in South-West London. J Hosp Infect 2020 Nov;106(3):621-625 [FREE Full text] [doi: 10.1016/j.jhin.2020.08.018] [Medline: 32841703]
4. Duy C, Nong VM, Van Ngo A, Doan TT, Nguyen TQ, Truong PT, et al. Nosocomial Coronavirus Disease Outbreak Containment, Hanoi, Vietnam, March-April 2020. Emerg Infect Dis 2021 Jan;27(1):10-17 [FREE Full text] [doi: 10.3201/eid2701.202656] [Medline: 33207153]
5. Abbas M, Robalo Nunes T, Martischang R, Zingg W, Iten A, Pittet D, et al. Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers. Antimicrob Resist Infect Control 2021 Jan 06;10(1):7 [FREE Full text] [doi: 10.1186/s13756-020-00875-7] [Medline: 33407833]
6. Zheng C, Hafezi-Bakhtiari N, Cooper V, Davidson H, Habibi M, Riley P, et al. Characteristics and transmission dynamics of COVID-19 in healthcare workers at a London teaching hospital. J Hosp Infect 2020 Oct;106(2):325-329 [FREE Full text] [doi: 10.1016/j.jhin.2020.07.025] [Medline: 32730771]

XSL•FO
RenderX

7. Verbeek VH, Rajamaki B, Ijaz S, Sauni R, Toomey E, Blackwood B, et al. Personal protective equipment for preventing highly infectious diseases due to exposure to contaminated body fluids in healthcare staff. Cochrane Database of Systematic Reviews 2020 Apr 15(4) [FREE Full text] [doi: 10.1002/14651858.CD011621.pub4]
8. Sud SR. COVID-19 and Keeping Clean: A Narrative Review To Ascertain the Efficacy of Personal Protective Equipment To Safeguard Health Care Workers Against SARS-CoV-2. Hosp Pediatr 2020 Jul 08;10(7):570-576. [doi: 10.1542/hpeds.2020-0135] [Medline: 32385055]
9. Bakken S. Telehealth: Simply a pandemic response or here to stay? J Am Med Inform Assoc 2020 Jul 01;27(7):989-990. [doi: 10.1093/jamia/ocaa132] [Medline: 32692845]
10. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
11. Vilendrer S, Patel B, Chadwick W, Hwa M, Asch S, Pageler N, et al. Rapid Deployment of Inpatient Telemedicine In Response to COVID-19 Across Three Health Systems. J Am Med Inform Assoc 2020 Jul 01;27(7):1102-1109 [FREE Full text] [doi: 10.1093/jamia/ocaa077] [Medline: 32495830]
12. Turer RW, Jones I, Rosenbloom S, Slovis C, Ward MJ. Electronic personal protective equipment: A strategy to protect emergency department providers in the age of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):967-971 [FREE Full text] [doi: 10.1093/jamia/ocaa048] [Medline: 32240303]
13. Overmann KM, Wu D, Xu C, Bindhu SS, Barrick L. Real-time locating systems to improve healthcare delivery: A systematic review. J Am Med Inform Assoc 2021 Jun 12;28(6):1308-1317. [doi: 10.1093/jamia/ocab026] [Medline: 33682009]
14. Li RC, Marafino BJ, Nielsen D, Baiocchi M, Shieh L. Assessment of a Real-Time Locator System to Identify Physician and Nurse Work Locations. JAMA Netw Open 2020 Feb 05;3(2):e1920352 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20352] [Medline: 32022876]
15. Hendrich A, Chow MP, Skierczynski BA, Lu Z. A 36-hospital time and motion study: how do medical-surgical nurses spend their time? Perm J 2008 Jun 01;12(3):25-34 [FREE Full text] [doi: 10.7812/tpp/08-021] [Medline: 21331207]
16. Ward DR, Ghali WA, Graham A, Lemaire JB. A real-time locating system observes physician time-motion patterns during walk-rounds: a pilot study. BMC Med Educ 2014 Feb 25;14(1):37 [FREE Full text] [doi: 10.1186/1472-6920-14-37] [Medline: 24568589]
17. D'Souza T, Rosen M, Bertram A, Apfel A, Desai SV, Garibaldi BT. Use of a Real-Time Location System to Understand Resident Location in an Academic Medical Center. J Grad Med Educ 2019 Jun;11(3):324-327 [FREE Full text] [doi: 10.4300/JGME-D-19-00026.1] [Medline: 31210865]
18. Jones TL, Schlegel C. Can real time location system technology (RTLS) provide useful estimates of time use by nursing personnel? Res Nurs Health 2014 Feb 11;37(1):75-84 [FREE Full text] [doi: 10.1002/nur.21578] [Medline: 24338915]
19. Berg B, Longley G, Dunitz J. Improving Clinic Operational Efficiency and Utilization with RTLS. J Med Syst 2019 Jan 30;43(3):56. [doi: 10.1007/s10916-019-1174-z] [Medline: 30701407]
20. Stahl JE, Drew MA, Leone D, Crowley RS. Measuring process change in primary care using real-time location systems: feasibility and the results of a natural experiment. Technol Health Care 2011 Nov 21;19(6):415-421. [doi: 10.3233/THC-2011-0638] [Medline: 22129942]
21. Vankipuram A, Traub S, Patel VL. A method for the analysis and visualization of clinical workflow in dynamic environments. J Biomed Inform 2018 Mar;79:20-31 [FREE Full text] [doi: 10.1016/j.jbi.2018.01.007] [Medline: 29410146]
22. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim W, Chow A. Use of a Real-Time Locating System for Contact Tracing of Health Care Workers During the COVID-19 Pandemic at an Infectious Disease Center in Singapore: Validation Study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
23. Venkataraman N, Poon BH, Siau C. Innovative use of health informatics to augment contact tracing during the COVID-19 pandemic in an acute hospital. J Am Med Inform Assoc 2020 Dec 09;27(12):1964-1967 [FREE Full text] [doi: 10.1093/jamia/ocaa184] [Medline: 32835358]
24. Fisher JA, Monahan T. Evaluation of real-time location systems in their hospital contexts. Int J Med Inform 2012 Oct;81(10):705-712. [doi: 10.1016/j.ijmedinf.2012.07.001] [Medline: 22857790]
25. Griffin JM, Hellmich TR, Pasupathy KS, Funni SA, Pagel SM, Srinivasan SS, et al. Attitudes and Behavior of Health Care Workers Before, During, and After Implementation of Real-Time Location System Technology. Mayo Clin Proc Innov Qual Outcomes 2020 Feb;4(1):90-98 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2019.10.007] [Medline: 32055774]
26. Sang AX, Tisdale RL, Nielsen D, Loica-Mersa S, Miller T, Chong I, et al. How Much Time are Physicians and Nurses Spending Together at the Patient Bedside? J Hosp Med 2019 Aug 01;14(8):468-473. [doi: 10.12788/jhm.3204] [Medline: 31112496]
27. Eby K. Coronavirus Timeline: Tracking major moments of COVID-19 pandemic in San Francisco Bay Area. ABC 7 News San Francisco. 2020. URL: https://abc7news.com/6047519/ [accessed 2020-06-01]
28. Ingram J. Medical School faculty member tests positive for coronavirus. The Stanford Daily. 2020 Mar 6. URL: https://www.stanforddaily.com/2020/03/06/medical-school-faculty-member-tests-positive-for-coronavirus/ [accessed 2020-06-01]
29. Wu D. Stanford Medicine is treating coronavirus patients. The Stanford Daily. 2020 Mar 5. URL: https://www.stanforddaily.com/2020/03/05/stanford-medicine-is-treating-coronavirus-patients-university-announces/ [accessed 2020-06-01]

XSL•FO
RenderX

30. Fernandez-Llatas C, Lizondo A, Monton E, Benedi J, Traver V. Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems. Sensors (Basel) 2015 Nov 30;15(12):29821-29840 [FREE Full text] [doi: 10.3390/s151229769] [Medline: 26633395]
31. Vankipuram M, Kahol K, Cohen T, Patel VL. Toward automated workflow analysis and visualization in clinical environments. J Biomed Inform 2011 Jun;44(3):432-440 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.015] [Medline: 20685315]
32. Hollingsworth JC, Chisholm CD, Giles BK, Cordell WH, Nelson DR. How Do Physicians and Nurses Spend Their Time in the Emergency Department? Annals of Emergency Medicine 1998 Jan;31(1):87-91. [doi: 10.1016/S0196-0644(98)70287-2]
33. Safaeinili N, Vilendrer S, Williamson E, Zhao Z, Brown-Johnson C, Asch SM, et al. Inpatient Telemedicine Implementation as an Infection Control Response to COVID-19: Qualitative Process Evaluation Study. JMIR Form Res 2021 Jun 16;5(6):e26452 [FREE Full text] [doi: 10.2196/26452] [Medline: 34033576]
34. Ribeira R, Shen S, Callagy P, Newberry J, Strehlow M, Quinn J. Telemedicine to Decrease Personal Protective Equipment Use and Protect Healthcare Workers. West J Emerg Med 2020 Sep 24;21(6):61-62 [FREE Full text] [doi: 10.5811/westjem.2020.8.47802] [Medline: 33052823]

Abbreviations

PPE: personal protective equipment
RT-PCR: reverse transcription–polymerase chain reaction
RTLS: real-time locating system
SHC: Stanford Health Care

Edited by G Eysenbach; submitted 31.03.21; peer-reviewed by C Margus, A Garg; comments to author 19.04.21; revised version received 27.05.21; accepted 13.06.21; published 26.07.21.

Please cite as:
Patel B, Vilendrer S, Kling SMR, Brown I, Ribeira R, Eisenberg M, Sharp C
Using a Real-Time Locating System to Evaluate the Impact of Telemedicine in an Emergency Department During COVID-19: Observational Study
J Med Internet Res 2021;23(7):e29240
URL: https://www.jmir.org/2021/7/e29240
doi:10.2196/29240
PMID:34236993

©Birju Patel, Stacie Vilendrer, Samantha M R Kling, Ian Brown, Ryan Ribeira, Matthew Eisenberg, Christopher Sharp. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Cloud-Based Telepathology System in China: Large-Sample Observational Study

Xianying He[1*], MSc; Linlin Wang[1*], MSc; Li Wang[2], MSc; Jinghong Gao[1,3], PhD; Fangfang Cui[1], MSc; Qianqian Ma[1], MSc; Wenjie Zhang[1,3], PhD; Lin Wang[1], MSc; Yunkai Zhai[4], PhD; Jie Zhao[1,3], PhD

[1]National Telemedicine Center of China, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[2]Department of Pathology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[3]National Engineering Laboratory for Internet Medical Systems and Applications, Zhengzhou, China

[4]School of Management Engineering, Zhengzhou University, Zhengzhou, China

[*]these authors contributed equally

Corresponding Author:
Jie Zhao, PhD
National Telemedicine Center of China
The First Affiliated Hospital of Zhengzhou University
1 Jianshe Road
Zhengzhou
China
Phone: 86 371 67966215
Fax: 86 371 67966215
Email: zhaojie@zzu.edu.cn

Abstract

Background: Whole-slide imaging allows the entire slide to be viewed in a manner that simulates microscopy; therefore, it is widely used in telepathology. However, managing the large digital files needed for whole-slide imaging is difficult. To solve this problem, we set up the Chinese National Cloud-Based Telepathology System (CNCTPS). CNCTPS has been running for more than 4 years and has accumulated a large amount of data.

Objective: The main purpose of this study was to comprehensively evaluate the effectiveness of the CNCTPS based on a large sample. The evaluation indicators included service volume, turnaround time, diagnosis accuracy, and economic benefits.

Methods: Details of 23,167 cases submitted to the CNCTPS from January 2016 to December 2019 were collected to analyze the service volume, turnaround time, and economic benefits. A total of 564 patients who visited the First Affiliated Hospital of Zhengzhou University and obtained final diagnoses were followed up to analyze the diagnostic accuracy of the CNCTPS.

Results: From 2016 to 2019, the service volume of the CNCTPS increased from 2335 to 9240, and the number of participating hospitals increased from 60 to 74. Consultation requests from county-level hospitals accounted for 86.57% (20,287/23,167). A total of 17,495 of 23,167 cases (75.52%) were confirmed, including 12,088 benign lesions, 5217 malignant lesions, and 190 borderline lesions. Of the cases, 3.85% (893/23,167) failed to be diagnosed for reasons such as poor slice quality and incomplete sampling. The median turnaround time was 16.93 hours and was shortened yearly (between 2018 and 2019: adjusted P=.01; other groups: adjusted P<.001); 82.88% cases were diagnosed in 48 hours. There was a discrepancy between the diagnosis and final diagnosis for 11 cases, including 4 false-positive cases and 7 false-negative cases. The sensitivity and specificity were 97.66% and 98.49%, respectively. The diagnostic accuracy of the system was 98.05%, with no statistical difference from the final diagnosis in the hospital (P=.55). By using this system, a total of US $300,000 was saved for patients every year.

Conclusions: The novel cloud-based telepathology system has the potential to relieve the shortage of pathologists in primary hospitals. It can also simultaneously reduce medical costs for patients in China. It should, therefore, be further promoted to enhance the efficiency, quantity, and quality of telepathology diagnoses.

(J Med Internet Res 2021;23(7):e23799) doi:10.2196/23799

KEYWORDS

telepathology; cloud-based system; whole-slide imaging; turnaround time; diagnostic accuracy; economic benefits

XSL•FO
RenderX

Introduction

Pathology diagnoses have been widely recognized as a gold standard for confirming diseases [1]. A precise and timely diagnosis is an indispensable precondition for further therapies [2]. However, there is a critical shortage and misdistribution of senior pathologists in resource-limited countries; China faces this challenge. According to statistics from the National Ministry of Health, there are 9841 licensed pathologists in China, but nearly 70% work in tertiary hospitals located in large cities [3]. Pathologists, especially senior and professional ones, are urgently needed in rural and remote areas [4-6]. To obtain a confirmed diagnosis and key guidance for subsequent therapies, undiagnosed pathology sections in county-level hospitals are usually mailed or personally transported to senior pathologists in tertiary hospitals. The procedure is complicated and costly. Furthermore, the valuable pathology sections are also at risk of being destroyed or lost.

Telepathology is a powerful tool that can be used to address this challenge by transmitting pathology images through telecommunication [7-9]. The first use of telepathology can be traced back to the 1960s in the United States, in which real-time black-and-white images were sent for interpretation. After a half a century of development, many uses of telepathology have been developed, with powerful features that can promptly transmit static, dynamic, and whole-slide images. The whole-slide imaging system is the most advanced means to view scanned and digitized slides in their entirety, with high-resolution digital images and superior zoom capability [10-12]. Whole-slide imaging has therefore been considered to be an ideal method for telepathology [13]. Despite considerable advancements, whole-slide imaging has several drawbacks [14], such as the need for large local storage space, network bandwidth constraints, cumbersome operation, occupied computing resources and large idle space, insufficient utilization rate, and difficulty in managing large digital files [15], which limits the application of whole-slide imaging.

To compensate for the shortcomings of whole-slide imaging, we established a Chinese National Cloud-Based Telepathology system (CNCTPS) based on an existing, mature telemedicine system of the National Telemedicine Center of China, with dual video and data drives, which solved the difficulty of telemedicine data interaction [16]. The CNCTPS was equipped with a deeply optimized storage model and analytical algorithm, which solved the problems of archiving classification and integration of large amounts of pathology data. This novel system can facilitate the prompt extraction and utilization of pathology data by doctors. The CNCTPS was deployed in December 2015; 83 hospitals were connected in total, making it the largest remote pathology network in China.

Previous studies have mainly focused on the construction and optimization of telepathology systems or analysis of the effect of system use, with a limited sample [17-22], and to date, there are no unified criteria to evaluate a telepathology system. Perron et al [23] evaluated diagnostic concordance and the turnaround time of a telepathology system. Chong et al [24] showed that telepathology shortened turnaround time and provided significant financial savings. Similarly, Zhou et al [6] reported service volume, turnaround time, and the concordance rate of a telepathology consultation service. However, these studies were mainly focused on one or some small and isolated aspects and did not comprehensively evaluate the service effect of the telepathology system. Thus, the aim of this study was to comprehensively evaluate the CNCTPS by evaluating 4 aspects—service volume, turnaround time, diagnosis accuracy, and economic benefits—which we chose after reviewing the literature on telepathology systems evaluation.

Methods

The Cloud-Based Telepathology System

Digitization of Pathology Sections

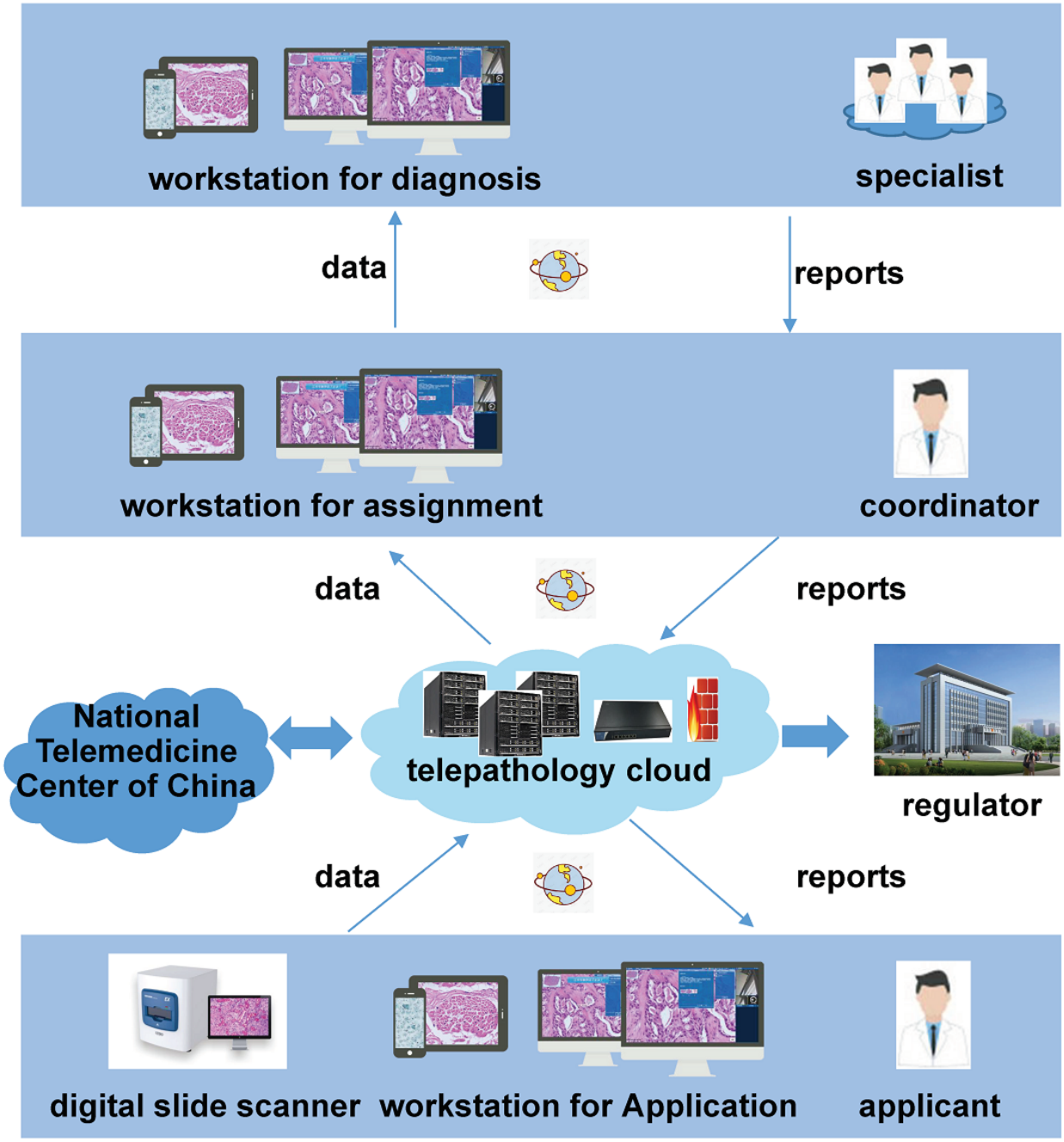
Participating hospitals were equipped with digital slide scanners and matched computer workstations (KF-PRO-005, Konfoong Biotech International Co Ltd), for converting traditional glass slices into whole-slide imaging. Whole-slide imaging of a slide could be completed within 40 seconds under a ×20 objective (0.47 μm/pixel) and within 100 seconds under a ×40 objective (0.5 μm/pixel). Scanning control software (K-Scanner 1.6.0.14, Konfoong Biotech International Co Ltd) and image browsing and management software (K-Viewer 1.5.3.1, Konfoong Biotech International Co Ltd) were used to control scanning and viewing in whole-slide imaging.

Data Storage and Transport

Whole-slide imaging and all other telepathology data were stored in dedicated servers located at the National Telemedicine Center of China to ensure the safety and speed of data storage, as well as the efficiency of data access by users. The overall design was based on a cloud-computing infrastructure service system, which was characterized by elastic expansion, high availability, and high stability. The system is equipped with wide-area and multilayer architecture, including access, application service, and data center layers (Figure 1). A private network with a bandwidth of up to 20 MB was used for data transmission.

XSL•FO
RenderX

Figure 1. Telepathology data storage and transmission.



Telepathology Management

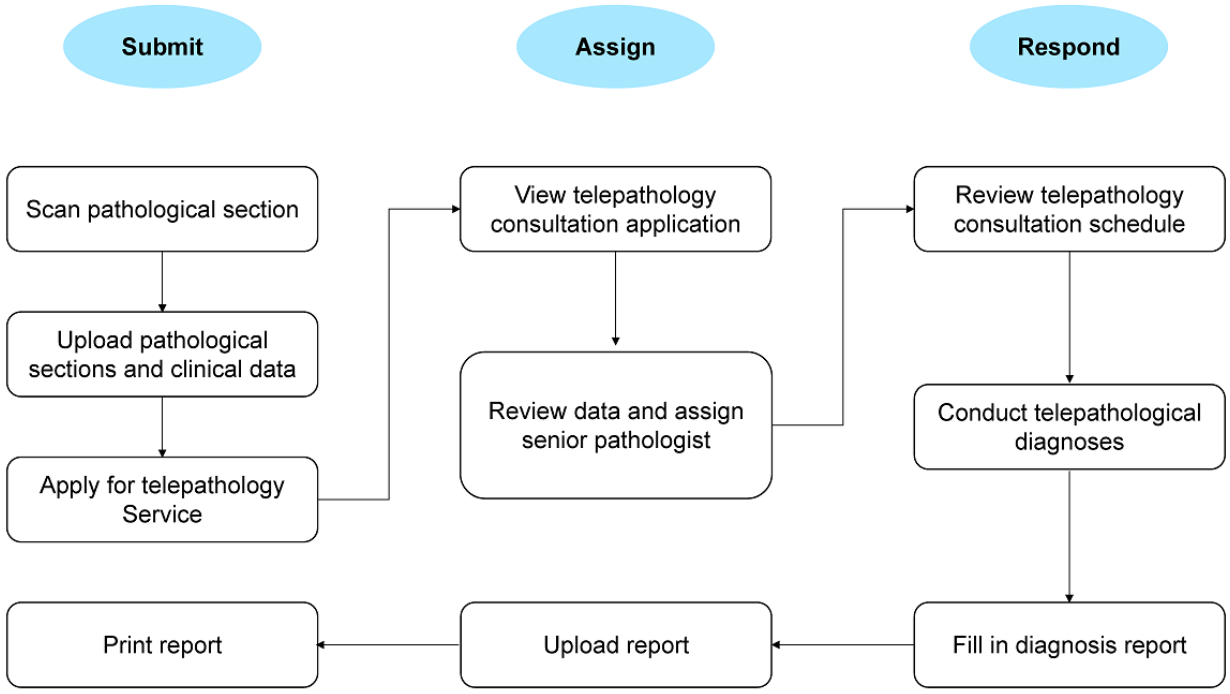
A web-based telepathology consultation system and mobile app were developed. Each has different functions for applicants, coordinators, and specialists. The web version was embedded in the telemedicine collaborative service platform of the National Telemedicine Center of China [25]; the app was independently developed and adapted for Android and iOS mobile phones and tablets (Multimedia Appendix 1).

Telepathology Consultation

There are 17 specialists from the Department of Pathology of the First Affiliated Hospital of Zhengzhou University who currently participate in telepathology consultation, including 8 professors and 9 associate professors specializing in different fields. The consultation is a voluntary activity with no charge. Pathologists from participating hospitals scanned and uploaded the slides to be diagnosed with patient information to the cloud platform. Coordinators from the National Telemedicine Center of China then assigned these cases to specialists (based on their specialties and fields), who are very likely to be able to provide confirmed diagnoses and valuable suggestions for corresponding therapies (Figure 2).

XSL•FO
RenderX

Figure 2. Telepathology consultation process.



CNCTPS Implementation Stages

The system was implemented in 3 stages. First, the participating hospitals were selected, starting in August 2015, based on medical service quality, readiness of their pathology departments and telemedicine services, and their willingness to use telepathology. Second, system hardware and software were deployed. Starting in January 2016, our technicians installed and debugged the equipment in participating hospitals. Third, personnel training and system maintenance were conducted. This included intensive training at the National Telemedicine Center of China (Multimedia Appendix 2) and on-site training in their hospitals. In addition, to ensure the normal operation of the system, technicians provide regular maintenance of the hardware and software in participating hospitals. System operation guides were also provided to the participating hospitals (Multimedia Appendix 3).

Data Collection

To analyze the service volume, turnaround time, and economic benefits of the CNCTPS, we collected all case data submitted from January 2016 to December 2019, which included demographic and clinical data, submitted hospital, case submission time, report issuance time, telepathology diagnosis, and specialist who made the diagnosis. After removing test cases, there were 23,167 cases. Specimens had been taken from multiple organs, which were divided into 26 groups.

To analyze the diagnostic accuracy of the CNCTPS, we followed up the final diagnosis of all the 23,167 cases through the hospital information system of the First Affiliated Hospital of Zhengzhou University. We searched and found that 564 cases had also been diagnosed directly in the First Affiliated Hospital of Zhengzhou University. The diagnostic accuracy of telepathology was calculated by using the final diagnosis in the First Affiliated Hospital of Zhengzhou University as the reference.

Statistical Analysis

Descriptive statistics were used to analyze characterize case data, including demographic characteristics of patients from whom samples were taken, diagnosis, histopathology type, and turnaround time. The median value and interquartile range are reported for continuous data, and percentages are reported for categorical data. The Kruskal–Wallis H test was used to compare turnaround time in different years, and the Nemenyi test was used for further multiple comparisons. The concordance between CNCTPS and final diagnoses was analyzed (complete concordance or variance with no clinical significance). The consistency was determined by the McNemar test and consistency check. All statistical analyses were performed using R software (version 4.0.0; R Foundation for Statistical Computing). All tests were 2-tailed, and $P<.05$ is considered statistically significant.

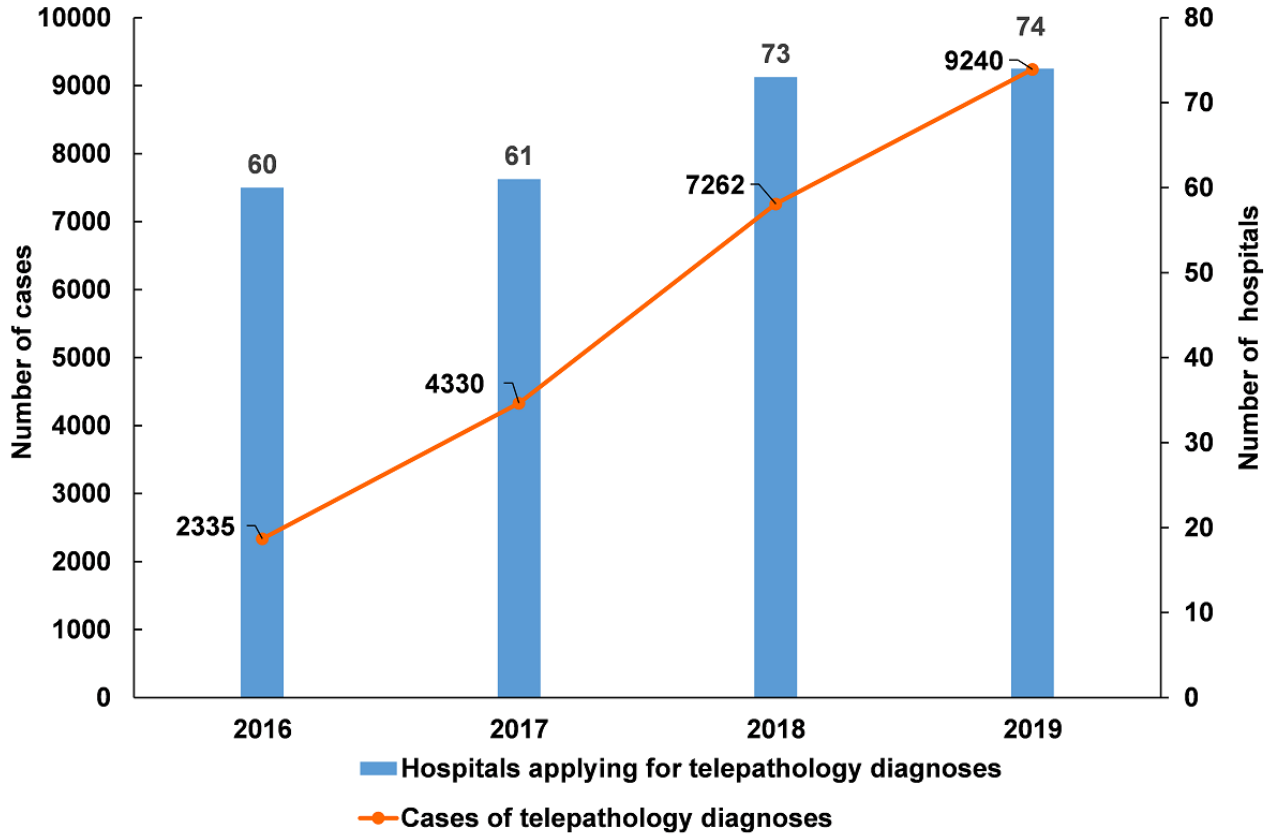
Results

CNCTPS Service Volume

During the 4-year study period from 2016 to 2019, 23,167 cases were submitted to the CNCTPS for consultation. The service volume of the CNCTPS was n=2335 in 2016; n=4330 in 2017; n=7262 in 2018; and n=9240 in 2019, with an average annual growth rate of 41.04%. A total of 83 hospitals participated in the telepathology consultation service. The number of participating hospitals also grew, from n=60 in 2016 to n=74 in 2019 (Figure 3).

XSL•FO
RenderX

Figure 3. Participating hospitals and submitted cases from 2016 to 2019.



Hospitals of different levels have joined the CNCTPS, including 17 city-level and 66 county-level hospitals. Among 2016 and 2019, city-level hospitals and county-level hospitals applied for 2880 (2880/23,167, 12.43%) and 20,287 (20,287/23,167, 87.57%) consultations, respectively. The number of county-level hospitals applying for consultation increased from n=49 in 2016 to n=63 in 2019, and the service volume also increased from n=2095 in 2016 to n=8317 in 2019. In city-level hospitals, the number of hospitals applying for consultations did not change, while the service volume showed an overall increasing trend (Table 1).

Table 1. Number of participating hospitals and service volume in different levels of hospitals from 2016 to 2019.

Hospital level	2016, n	2017, n	2018, n	2019, n
City-level				
Hospitals	11	9	14	11
Service volume	240	657	1060	923
County-level				
Hospitals	49	52	59	63
Service volume	2095	3673	6202	8317

Characteristics of Cases Submitted to the CNCTPS

The locations, from which specimens had been taken, were divided into 26 groups (Table 2).

Of the 23,167 patients represented by case data, 9519 (41.09%) were male and 13,648 (58.91%) were female (Table 3). The median age of the patients, from whom specimens were taken, was 53 years (mean 52.86 years, range 1 day to 98 years). There were 17,495 out of 23,167 cases (75.52%) with confirmed diagnoses; 4779 out of 23,167 cases (20.63%) needed further examination, and most (4007/4779, 83.85%) required immunohistochemical examination. The other 893 (893/23,167, 3.85%) cases failed to be diagnosed, and poor slice quality and incomplete sampling were the main reasons thereof.

Among 17,495 confirmed cases, 12,088 were benign lesions, 5217 were malignant lesions, and 190 were borderline lesions. In total, 52.18% (12,088/23,167) benign cases and 22.52% (5217/23,167) malignant cases had been confirmed. The proportion of malignant lesions in the esophagus, lung/mediastinum, urinary, and thoracic cavity/pleura was higher than that of benign lesions (Figure 4). In the other 22 tissue types, the proportion of benign lesions was higher than that of malignant lesions.

XSL•FO
RenderX

Table 2. Anatomic sites of specimens.

Anatomic site	Value (n=23,167), n (%)
Uterus	4074 (17.59)
Gastrointestinal	3643 (15.72)
Bone and soft tissue	2900 (12.52)
Breast	1488 (6.42)
Esophagus	1181 (5.10)
Lung/mediastinum	1175 (5.07)
Thyroid	1065 (4.60)
Head and neck	963 (4.16)
Female genital organs except for uterus	854 (3.69)
Oral cavity	850 (3.67)
Urinary	730 (3.15)
Male genital organs	725 (3.13)
Hepatobiliary and pancreas	647 (2.79)
Respiratory tract	595 (2.57)
Eyes and ears	467 (2.02)
Skin	422 (1.82)
Lymphoid organs	316 (1.36)
Miscellaneous	262 (1.13)
Hydrothorax/ascites	168 (0.73)
Central nervous system	145 (0.63)
Anus and perianal	116 (0.50)
Abdominal cavity/peritoneum/postperitoneum	114 (0.49)
Others	110 (0.47)
Pelvic cavity	70 (0.30)
Adrenal glands	46 (0.20)
Thoracic cavity/ pleura	41 (0.18)

XSL•FO
RenderX

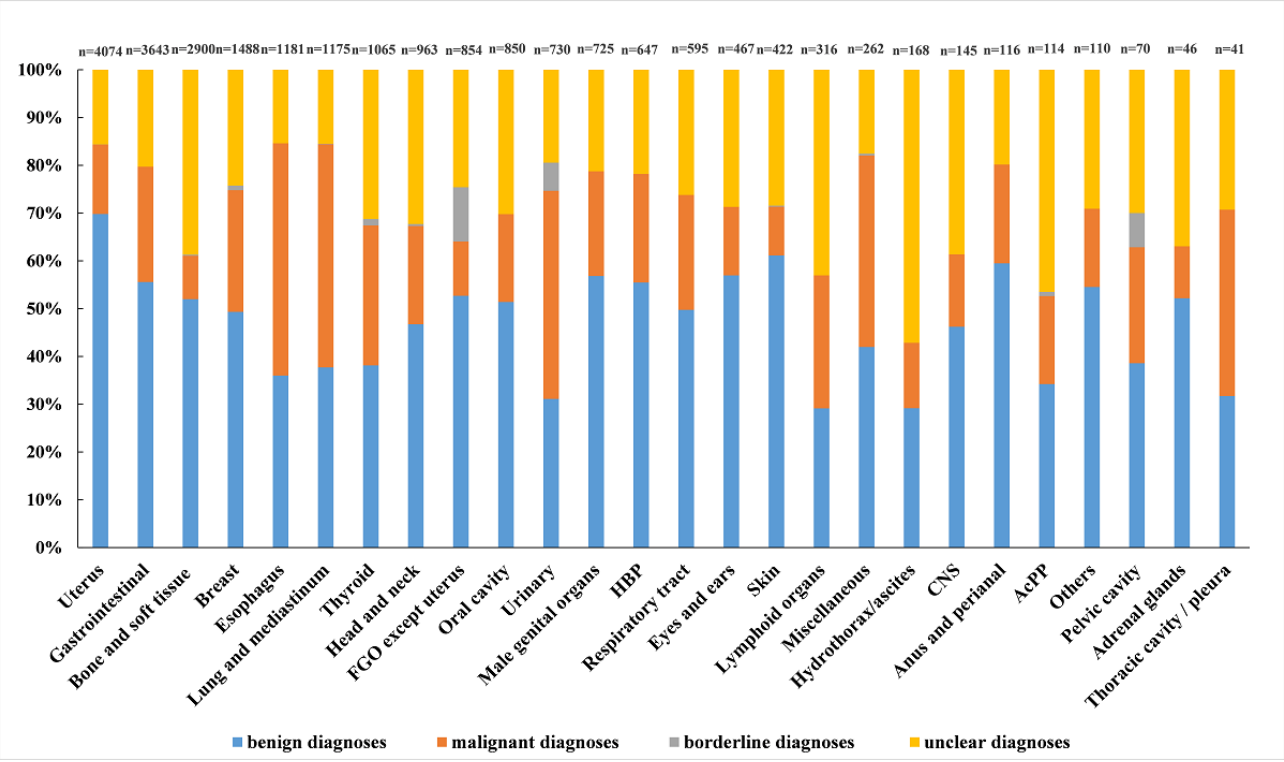
Table 3. Case and patient characteristics.

Variables	Values
Sex, n (%)	
Male	9519 (41.09)
Female	13,648 (58.91)
Age, mean (range)	52.86 (1 day to 98 years)
Diagnosis types, n (%)	
Confirmed	17,495 (75.52)
Needed further examination[a]	4779 (20.63)
Failed to be diagnosed[b]	893 (3.85)
Histopathology types, n (%)	
Benign	12,088 (52.18)
Borderline	190 (0.82)
Malignant	5217 (22.52)
Unclear	5672 (24.48)

[a]Further examinations included immunohistochemistry assay (4007/4779, 83.85%), clinical examinations (675/4779, 14.12%), gene detection (88/4779, 1.84%), and special staining (9/4779, 0.19%).

[b]The reasons included poor slice quality (531/893, 59.46%), incomplete sampling (336/893, 37.63%) and intractable cases (26/893, 2.91%).

Figure 4. Histopathology distribution for 26 anatomic locations. AcPP: abdominal cavity/peritoneum/postperitoneum; CNS: central nervous system; HBP: hepatobiliary and pancreas; FGO: female genital organs.



CNCTPS Turnaround Time

The turnaround time, the time from transmitting whole-slide images to the issuance of diagnostic reports, was a median of 16.93 hours (IQR 32.59; mean 24.93 hours, range 100 seconds to 167.97 hours). Experts' opinion reports were released within 12 hours in 10,244 of the 23,167 cases (10,244/23,167, 44.05%) and within 72 hours in 21,286 cases (21,286/23,167, 91.88%) (Table 4).

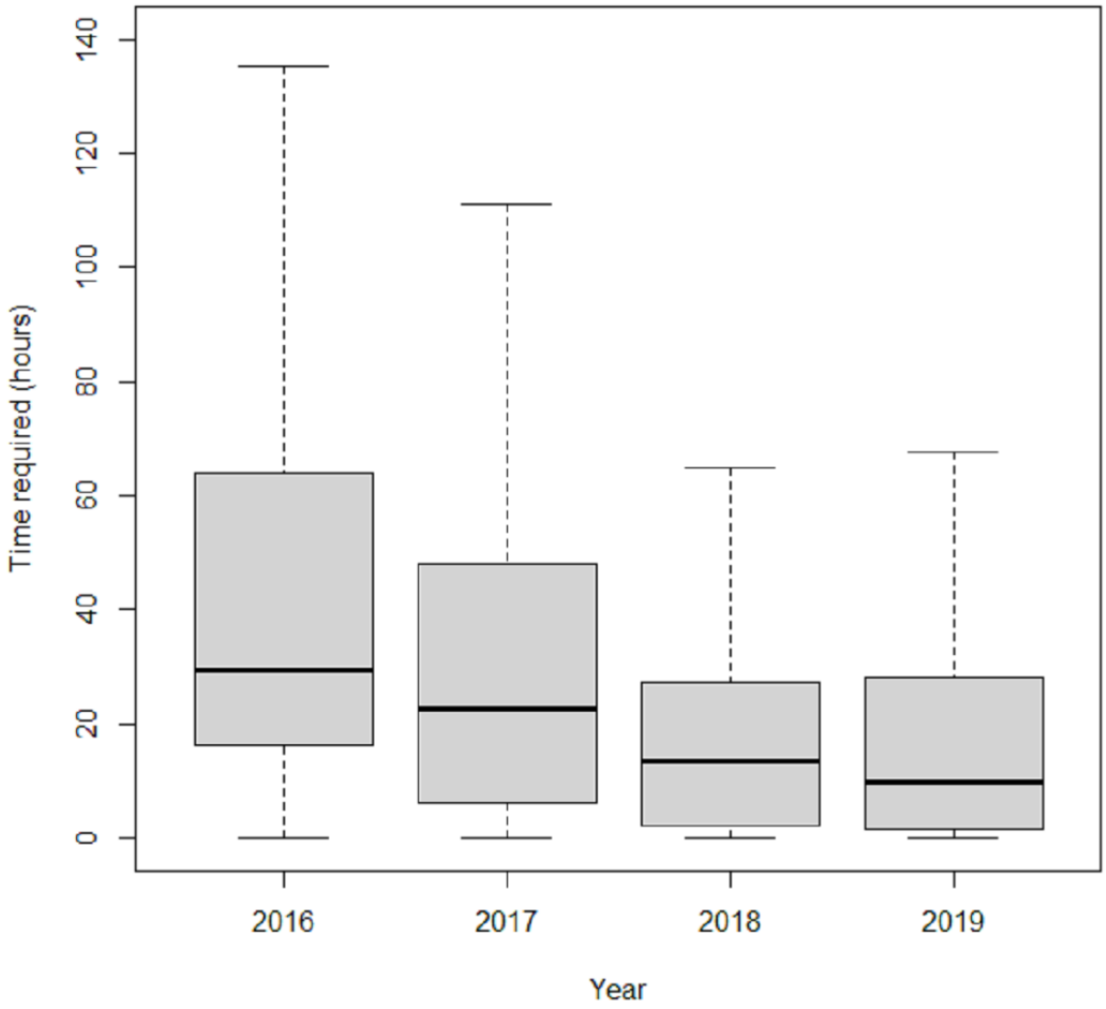
XSL•FO
RenderX

Table 4. Turnaround time for expert reports.

Time required (hours)	Cases, n (%)	Cumulative %
Time≤12	10,204 (44.05)	44.05
12<time≤24	4631 (19.99)	64.04
24<time≤48	4366 (18.85)	82.88
48<time≤72	2085 (9.00)	91.88
Time>72	1881 (8.12)	100

The difference in distribution of turnaround time in different years (Figure 5), was statistically significant (H=1433.62, *P*<.001). The median turnaround time gradually decreased in turn, from 29.36 hours in 2016 to 9.75 hours in 2019, and differences between subsequent years were statistically significant with pairwise comparison (between 2018 and 2019 adjusted *P*=.01, other groups adjusted *P*<.001).

Figure 5. Turnaround time distribution.



CNCTPS Diagnostic Accuracy

Of 564 diagnosed by both the CNCTPS and pathologists in the hospital, 553 cases diagnosed by the CNCTPS were consistent with the final diagnosis made by pathologists in hospital; that is, the accuracy rate was 98.05%. In the other 11 cases—4 false-positive cases and 7 false-negative cases—5 of the 11 cases occurred in the uterus (Table 5).

The sensitivity and specificity were 97.66% and 98.49%, respectively (Table 6). The Youden index was 0.96. The positive and negative predictive values were 98.65% and 97.39%, respectively. No statistical difference was observed between telepathology diagnosis and final diagnosis (*P*=.55), which showed good consistency (κ=0.96, *P*<.001).

XSL•FO
RenderX

Table 5. Discordant cases (between telepathology and final diagnoses).

Type	Sample source	Telepathology diagnosis	Final diagnosis	Annotation
False positive[a]				
Case 1	Lung/mediastinum	Poorly differentiated carcinoma	Immunoglobulin M–positive lymphoproliferative disease with alveolar epithelial atypical hyperplasia	Unlabeled immunohistochemistry results
Case 2	Respiratory tract	Highly differentiated squamous cell carcinoma	Squamous papillary hyperplasia with local typical hyperplasia	Intractable case
Case 3	Uterus	Endometrial complex hyperplasia, local atypical hyperplasia, focal canceration	Simple endometrial hyperplasia	N/A[b]
Case 4	Uterus	Papillary squamous cell carcinoma	High-grade SIL[c] involving glands	A bleeding background on the section
False negative[d]				
Case 1	Female genital organs	Vulva: chronic inflammation with low-grade SIL Cervix: chronic inflammation with high-grade SIL	Vulva: highly differentiated squamous cell carcinoma with local superficial infiltration (depth of infiltration <1 mm) Cervix: chronic cervicitis, focal high-grade SIL, and involving glands	Unlabeled immunohistochemistry results
Case 2	Skin	Chronic inflammation, squamous epithelial hyperplasia with hyperkeratosis and parakeratosis	Superficial spreading malignant melanoma	Intractable case
Case 3	Uterus	Chronic inflammation, glandular hyperplasia	Minimal deviation adenocarcinoma	N/A
Case 4	Uterus	Chronic cervicitis with focal high-grade SIL and involving glands	High-grade SIL involving glands and squamous cell carcinoma in situ	N/A
Case 5	Thyroid	Adenomatous nodular goiter with fibrosis and chronic lymphocytic thyroiditis around	Follicular carcinoma	Unlabeled immunohistochemistry results
Case 6	Urinary	Mucosal polypoid hyperplasia with atypical urothelial hyperplasia	High-grade urothelial carcinoma	N/A
Case 7	Uterus	SIL	Squamous cell carcinoma	A bleeding background on the section

[a]Cases that were malignant in telepathology diagnosis but benign in final diagnosis were considered false positive.

[b]N/A: not applicable.

[c]SIL: squamous intraepithelial lesion.

[d]Cases that were malignant in the final diagnosis but benign in telepathology diagnosis were considered false negative.

Table 6. Validity of CNCTPS diagnoses.

Telepathology diagnosis	Final diagnosis		
	Positive, n	Negative, n	Total, n
Positive	292	4	296
Negative	7	261	268
Total	299	265	564

Economic Benefits of the CNCTPS

Telepathology consultation is free and avoids the need for patients having to visit higher-level hospitals. Therefore, consultation and travel costs were saved. At the same time, food costs were lower in the local area. Thus, compared with the traditional pathology consultation, diagnosis via the CNCTPS results in cost-savings of 378.5 RMB (approximately US $50) per patient (Table 7). In terms of the annual telepathology consultation cases, the total amount is substantial—approximately $300,000 per year.

XSL•FO
RenderX

Table 7. Cost savings for each patient applied for telepathology consultation.

Type of costs	Telepathology consultation cost (RMB[a])	Traditional pathology consultation cost (RMB)	Costs savings (RMB)
Consultation costs	0	148.5[b]	148.5
Travel costs[c]	0	200	200
Food costs[c]	25	55	30
Total	25	403.5	378.5

[a]RMB: Renminbi; an approximate exchange rate of 6.46 RMB = US $1 is applicable.

[b]Traditional consultation costs referred to the pathology consultation charges in Henan Province.

[c]For each patient, one person going for a consultation was assumed. Travel and food costs were calculated using estimates of local corresponding average expenses.

Discussion

Principal Results

The cloud-based system can quickly process data with large memory requirements, thereby overcoming the difficulties of large whole-slide imaging file management. This study reported on one of the largest cloud-based telepathology systems in China and evaluated its operation results. This study used a large sample size, which provides an in-depth practical understanding of the cloud-based telepathology system in China, and gives suggestions for further evaluations and improvements of the telepathology system. This system served 23,167 cases from 2016 to 2019. The median turnaround time was 16.93 hours, which decreased from 29.36 hours in 2016 to 9.75 hours in 2019. The diagnostic accuracy was 98.05%, and approximately $300,000 were collectively saved by patients each year. The CNCTPS has proven to be highly reliable and plays an important role in facilitating the distribution of limited senior pathologist resources in China.

A total of 83 hospitals are covered by the CNCTPS, which is the largest telepathology network in China. Compared with the 6, 24, and 60 workstations in other reported telepathology networks [23,24,26], the CNCTPS covers more medical institutions. Case data for more than 20,000 patients were diagnosed by the CNCTPS in 4 years. To the best of our knowledge, this is the largest sample size in a study on telepathology system use and operation and is much higher than those in similar literature [24,26-28]. The amount of case data reviewed and the number of participating hospitals increased each year, consistent with findings reported by Chen et al [26] and Zhou et al [6]. Most case data (20,287/23,167, 87.57%) had been submitted by county-level hospitals because the shortage of pathologists in China's county-level hospitals is more severe than that in city-level hospitals.

A total of 893 cases failed to be diagnosed by the CNCTPS, of which only 26 cases were complicated enough that needed to be consulted in a higher-level hospital, while others were due to incomplete sampling and poor slice quality. Standard materials and good slice preparation are the main factors affecting telepathology diagnosis [29], which require experienced pathology technicians. Although we had conducted theory and practical operation training for pathology technicians in the early stage of CNCTPS construction, incomplete sampling and poor slice quality were still the main reasons for failed diagnoses. Strengthening the training of telepathology staff in the later stages of system operation is still needed. In terms of histopathology type, an analysis of ten-year telepathology cases in Tanzania showed a higher proportion of benign cases, which reported 56.1% benign and 40.8% malignant diseases [27]. We reached a similar conclusion: the proportion of benign cases (12,088/23,167, 52.18%) was higher than that of malignant cases (5217/23,167, 22.52%).

The average turnaround time of the CNCTPS was 24.93 hours, which is shorter than the 38 hours reported by Zhou et al [6] and 66 hours reported by Völker et al [27] but is slightly higher than the 0.7 days (ie, 16.8 hours) reported by Chong et al [24]. The majority (14,835/23,167, 64.04%) of cases were diagnosed within 24 hours, which is higher than the 61.5% reported by Chen et al [26] and slightly lower than the 64% reported by Perron et al [23], but the proportion within 48 hours (82.88% vs 70.00%) and 72 hours (91.88% vs 80.00%) were higher than those reported by Perron et al [23]. Nonetheless, the median turnaround time decreased annually during the 4 years, indicating that the CNCTPS operates well.

Compared to static images in the early stage of telepathology, whole-slide imaging allows the entire slide to be viewed in a manner that simulates microscopy [13]. A recent meta-analysis [30] shows that the weighted mean of the concordance rates between telepathology and conventional microscopy was 91.1% up to 2000, and from 2000 onward, the weighted mean of the concordance rates was 97.2%. It has been asserted that the reasons for increased consistency rate in recent years should be attributed to the increased use of whole-slide imaging [30]. The range of diagnostic concordance rates between whole-slide imaging and traditional electron microscopy is 89% to 100% [31-35], and the average value is 96.9%. Our study demonstrated similar results (98.05%). Moreover, no statistically significant differences were found (P=.55) between whole-slide imaging and traditional pathology diagnosis, and the consistency of diagnostic results was excellent, which further confirmed the accuracy of whole-slide imaging.

Some cost-effectiveness studies have demonstrated that telemedicine can reduce costs [36], but not all [37,38]. Cost-utility and cost-effectiveness studies for telepathology are rare. Meléndez-Álvarez et al [17] only evaluated the cost of their telepathology system, which saved US $410. Vosoughi et al [39] evaluated the cost-efficiency of their real-time nonrobotic

XSL•FO
RenderX

telepathology system, which saved US $10,767.10 per year. In our study, cost savings for patients were estimated. During the 4 years of the telepathology system operation, approximately US $300,000 per year was saved by patients.

Limitations

To the best of our knowledge, this is the first study to comprehensively evaluate the operation of a telepathology system based on a large sample. The CNCTPS showed fast responsiveness and high accuracy. However, owing to the limited information collected by the CNCTPS, this study did not analyze the reasons for cases with long turnaround time or the reasons for false positives and false negatives. In addition, only the costs saved for patients were evaluated in the economic benefits of the CNCTPS. The economic impact of telemedicine is a collaborative and complex process in which different economic, social, and political actors can be involved [38], and the construction of our system is a public welfare project initiated by the government and a powerful hospital. Most of the digital slide scanners were donated to the participating hospitals, and the private network was free.

Future Work

Turnaround time and diagnostic accuracy are the main criteria used to evaluate a telepathology system, and further work is required to explore the factors that influence turnaround time and diagnostic accuracy. First, it is necessary to analyze the causes of cases with long turnaround time through a survey of pathologists, especially those with turnaround times longer than 72 hours. Second, more investigation for incorrectly diagnosed images is needed. In addition, adding a follow-up module to the CNCTPS is necessary to allow the final diagnosis result of each case to be easily followed up. Finally, a user satisfaction survey should be conducted, with thorough questionnaires or in-depth interviews, in a subsequent study to improve the system.

Conclusions

The CNCTPS has proven to be highly reliable. It can provide rapid telepathology diagnoses to participating hospitals that are consistent with the final diagnosis. The application of this system reduces financial costs and time for patients, facilitating the distribution of limited senior pathologist resources in China. Therefore, we believe telepathology services will become more widespread, in more regions worldwide, especially those with insufficient medical resources.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant number 2017YFC0909901).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The web and app versions of the cloud-based telepathology system.
[PPTX File , 23268 KB - jmir_v23i7e23799_app1.pptx]

Multimedia Appendix 2
Training documents.
[PDF File (Adobe PDF File), 13521 KB - jmir_v23i7e23799_app2.pdf]

Multimedia Appendix 3
System operation guide for the cloud-based telepathology system.
[PDF File (Adobe PDF File), 6616 KB - jmir_v23i7e23799_app3.pdf]

References

1. Qiu M, Qiu H, Jin Y, Wei X, Zhou Y, Wang Z, et al. Pathologic diagnosis of pancreatic adenocarcinoma in the United States: its status and prognostic value. J Cancer 2016;7(6):694-701 [FREE Full text] [doi: 10.7150/jca.14185] [Medline: 27076851]
2. Tysarowski A, Nasierowska-Guttmejer A. Quality and practical aspects of pathological and molecular diagnostics in metastatic colorectal cancer (mCRC). Contemp Oncol (Pozn) 2018;22(2):75-85 [FREE Full text] [doi: 10.5114/wo.2018.77047] [Medline: 30150883]
3. Huang Y, Lei Y, Wang Q, Li D, Ma L, Guo L, et al. Telepathology consultation for frozen section diagnosis in China. Diagn Pathol 2018 May 14;13(1):29 [FREE Full text] [doi: 10.1186/s13000-018-0705-0] [Medline: 29759085]
4. Li X, Gong E, McNutt MA, Liu J, Li F, Li T, et al. Assessment of diagnostic accuracy and feasibility of dynamic telepathology in China. Hum Pathol 2008 Feb;39(2):236-242. [doi: 10.1016/j.humpath.2007.06.008] [Medline: 17950781]

XSL•FO
RenderX

5. Gimbel DC, Sohani AR, Prasad Busarla SV, Kirimi JM, Sayed S, Okiro P, et al. A static-image telepathology system for dermatopathology consultation in East Africa: the Massachusetts General Hospital Experience. J Am Acad Dermatol 2012 Nov;67(5):997-1007. [doi: 10.1016/j.jaad.2011.12.036] [Medline: 22341607]
6. Zhou C, Rahemtulla A, Hua LT, Shi H. Telepathology consultation in China using whole slide image and an internet based platform. Diagn Pathol 2013 Sep 30;8(S1):1-3. [doi: 10.1186/1746-1596-8-s1-s10]
7. Leong FJ. Practical applications of Internet resources for cost-effective telepathology practice. Pathology 2001 Nov;33(4):498-503. [doi: 10.1080/00313020127147] [Medline: 11827419]
8. Weinstein RS, Descour MR, Liang C, Bhattacharyya AK, Graham AR, Davis JR, et al. Telepathology overview: from concept to implementation. Hum Pathol 2001 Dec;32(12):1283-1299. [doi: 10.1053/hupa.2001.29643] [Medline: 11774159]
9. Farahani N, Pantanowitz L. Overview of telepathology. Clin Lab Med 2016 Mar;36(1):101-112. [doi: 10.1016/j.cll.2015.09.010] [Medline: 26851668]
10. Banavar SR, Chippagiri P, Pandurangappa R, Annavajjula S, Rajashekaraiah PB. Image montaging for creating a virtual pathology slide: an innovative and economical tool to obtain a whole slide image. Anal Cell Pathol (Amst) 2016;2016:9084909 [FREE Full text] [doi: 10.1155/2016/9084909] [Medline: 27747147]
11. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019 May;20(5):e253-e261. [doi: 10.1016/S1470-2045(19)30154-8] [Medline: 31044723]
12. Costello SSP, Johnston DJ, Dervan PA, O'Shea DG. Development and evaluation of the virtual pathology slide: a new tool in telepathology. J Med Internet Res 2003;5(2):e11 [FREE Full text] [doi: 10.2196/jmir.5.2.e11] [Medline: 12857667]
13. Pantanowitz L, Dickinson K, Evans AJ, Hassell LA, Henricks WH, Lennerz JK, et al. ATA clinical guidelines for telepathology. Telemed J E Health 2014 Nov;20(11):1049-1056. [doi: 10.1089/tmj.2014.9976] [Medline: 25384254]
14. Khalbuss WE, Pantanowitz L, Parwani AV. Digital imaging in cytopathology. Patholog Res Int 2011;2011:264683 [FREE Full text] [doi: 10.4061/2011/264683] [Medline: 21785680]
15. French JMR, Betney DT, Abah U, Bhatt N, Internullo E, Casali G, et al. Digital pathology is a practical alternative to on-site intraoperative frozen section diagnosis in thoracic surgery. Histopathology 2019 May;74(6):902-907. [doi: 10.1111/his.13804] [Medline: 30537290]
16. Zhai Y, Gao J, Chen B, Shi J, Wang L, He X, et al. Design and application of a telemedicine system jointly driven by videoconferencing and data exchange: practical experience from Henan Province, China. Telemed J E Health 2020 Jan;26(1):89-100. [doi: 10.1089/tmj.2018.0240] [Medline: 31050599]
17. Meléndez-Álvarez B, Robayo O, Gil-Guillén V, Carratalá-Munuera M. Design and validation of a low-cost telepathology system. Telemed J E Health 2017 Dec;23(12):976-982. [doi: 10.1089/tmj.2017.0030] [Medline: 28537789]
18. Hajesmaeel Gohari S, Bahaadinbeigy K, Malekpour Afshar R, Asadi-Shekaari M. A trial study of static telepathology in Iran. Health Care Manag (Frederick) 2018;37(3):262-267. [doi: 10.1097/HCM.0000000000000227] [Medline: 29957660]
19. Kldiashvili E, Schrader T, Burduli A, Ghortlishvili G. Application of medical information system for telepathology--Georgian experience. Telemed J E Health 2010;16(6):699-704. [doi: 10.1089/tmj.2010.0022] [Medline: 20618089]
20. Siegel G, Regelman D, Maronpot R, Rosenstock M, Nyska A. New technologies: real-time telepathology systems-novel cost-effective tools for real-time consultation and data sharing. Toxicol Pathol 2017 Dec;45(8):1039-1042. [doi: 10.1177/0192623317743722] [Medline: 29233077]
21. Goswami R, Pi D, Pal J, Cheng K, Hudoba De Badyn M. Performance evaluation of a dynamic telepathology system (Panoptiq™) in the morphologic assessment of peripheral blood film abnormalities. Int J Lab Hematol 2015 Jun;37(3):365-371. [doi: 10.1111/ijlh.12294] [Medline: 25227659]
22. Kadaba V, Ly T, Noor S, Chhut SV, Hinsch N, Stauch G, et al. A hybrid approach to telepathology in Cambodia. J Telemed Telecare 2013 Dec;19(8):475-478. [doi: 10.1177/1357633X13512071] [Medline: 24197402]
23. Perron E, Louahlia S, Nadeau L, Boilard F, Ing M, Orain M, et al. Telepathology for intraoperative consultations and expert opinions: the experience of the Eastern Québec Telepathology Network. Arch Pathol Lab Med 2014 Sep;138(9):1223-1228 [FREE Full text] [doi: 10.5858/arpa.2013-0466-OA] [Medline: 25171705]
24. Chong T, Palma-Diaz MF, Fisher C, Gui D, Ostrzega NL, Sempa G, et al. The California Telepathology Service: UCLA's experience in deploying a regional digital pathology subspecialty consultation network. J Pathol Inform 2019;10:31 [FREE Full text] [doi: 10.4103/jpi.jpi_22_19] [Medline: 31620310]
25. National Telemedicine Center of China. Telemedicine collaboration. 2014. URL: http://www.htcc.org.cn/page/telem-colla.html [accessed 2021-06-30]
26. Chen J, Jiao Y, Lu C, Zhou J, Zhang Z, Zhou C. A nationwide telepathology consultation and quality control program in China: implementation and result analysis. Diagn Pathol 2014 Dec;9 Suppl 1:S2 [FREE Full text] [doi: 10.1186/1746-1596-9-S1-S2] [Medline: 25565398]
27. Völker H, Müller-Hermelink H, Stüfe A, Strehl A, Pötzl L, Stauch G. [Ten years of telepathology for a mission hospital in Tanzania]. Pathologe 2019 Sep;40(5):519-526. [doi: 10.1007/s00292-019-0641-0] [Medline: 31338564]
28. Têtu B, Perron, Louahlia S, Paré G, Trudel M, Meyer J. The Eastern Québec Telepathology Network: a three-year experience of clinical diagnostic services. Diagn Pathol 2014;9 Suppl 1:S1. [doi: 10.1186/1746-1596-9-S1-S1] [Medline: 25564940]

XSL•FO
RenderX

29. Montgomery ND, Tomoka T, Krysiak R, Powers E, Mulenga M, Kampani C, et al. Practical successes in telepathology experiences in Africa. Clin Lab Med 2018 Mar;38(1):141-150 [FREE Full text] [doi: 10.1016/j.cll.2017.10.011] [Medline: 29412878]
30. Dietz RL, Hartman DJ, Pantanowitz L. Systematic review of the use of telepathology during intraoperative consultation. Am J Clin Pathol 2020 Jan 02;153(2):198-209. [doi: 10.1093/ajcp/aqz155] [Medline: 31618416]
31. Ramey J, Fung KM, Hassell LA. Use of mobile high-resolution device for remote frozen section evaluation of whole slide images. J Pathol Inform 2011;2:41. [doi: 10.4103/2153-3539.84276] [Medline: 21969922]
32. Bauer TW, Slaw RJ. Validating whole-slide imaging for consultation diagnoses in surgical pathology. Arch Pathol Lab Med 2014 Nov;138(11):1459-1465. [doi: 10.5858/arpa.2013-0541-OA] [Medline: 24840034]
33. Pradhan D, Monaco S, Parwani A, Ahmed I, Duboy J, Pantanowitz L. Evaluation of panoramic digital images using Panoptiq for frozen section diagnosis. J Pathol Inform 2016;7(1):26. [doi: 10.4103/2153-3539.181770] [Medline: 27217976]
34. Tabata K, Mori I, Sasaki T, Itoh T, Shiraishi T, Yoshimi N, et al. Whole-slide imaging at primary pathological diagnosis: validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. Pathol Int 2017 Nov;67(11):547-554. [doi: 10.1111/pin.12590] [Medline: 28980740]
35. Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Mod Pathol 2019 Jul;32(7):916-928 [FREE Full text] [doi: 10.1038/s41379-019-0205-0] [Medline: 30778169]
36. Müller KI, Alstadhaug KB, Bekkelund SI. Acceptability, feasibility, and cost of telemedicine for nonacute headaches: a randomized study comparing video and traditional consultations. J Med Internet Res 2016 May 30;18(5):e140 [FREE Full text] [doi: 10.2196/jmir.5221] [Medline: 27241876]
37. Richter KP, Shireman TI, Ellerbeck EF, Cupertino AP, Catley D, Cox LS, et al. Comparative and cost effectiveness of telemedicine versus telephone counseling for smoking cessation. J Med Internet Res 2015 May 08;17(5):e113 [FREE Full text] [doi: 10.2196/jmir.3975] [Medline: 25956257]
38. de la Torre-Diez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
39. Vosoughi A, Smith PT, Zeitouni JA, Sodeman GM, Jorda M, Gomez-Fernandez C, et al. Frozen section evaluation via dynamic real-time nonrobotic telepathology system in a university cancer center by resident/faculty cooperation team. Hum Pathol 2018 Aug;78:144-150. [doi: 10.1016/j.humpath.2018.04.012] [Medline: 29723604]

Abbreviations

CNCTPS: Chinese National Cloud-Based Telepathology System
RMB: Renminbi
SIL: squamous intraepithelial lesion

Edited by R Kukafka; submitted 24.08.20; peer-reviewed by S Rostam Niakan Kalhori, P Fontelo; comments to author 22.09.20; revised version received 18.12.20; accepted 24.05.21; published 29.07.21.

Please cite as:
He X, Wang L, Wang L, Gao J, Cui F, Ma Q, Zhang W, Wang L, Zhai Y, Zhao J
Effectiveness of a Cloud-Based Telepathology System in China: Large-Sample Observational Study
J Med Internet Res 2021;23(7):e23799
URL: https://www.jmir.org/2021/7/e23799
doi:10.2196/23799
PMID:34326037

©Xianying He, Linlin Wang, Li Wang, Jinghong Gao, Fangfang Cui, Qianqian Ma, Wenjie Zhang, Lin Wang, Yunkai Zhai, Jie Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Technology Use and BMI: Evidence From a Cross-sectional Analysis of an Adolescent Cohort Study

Chen Shen[1,2], PhD; Iroise Dumontheil[3], PhD; Michael Thomas[3], PhD; Martin Röösli[4,5], PhD; Paul Elliott[1,2,6], PhD; Mireille Toledano[1,2,7], PhD

[1]MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

[2]National Institute for Health Research Health Protection Research Unit in Chemical and Radiation Threats and Hazards, Imperial College London, London, United Kingdom

[3]Department of Psychological Sciences, Birkbeck, University of London, London, United Kingdom

[4]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

[5]University of Basel, Basel, Switzerland

[6]NIHR Biomedical Research Centre, Imperial College London, London, United Kingdom

[7]Mohn Centre for Children's Health and Wellbeing, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Mireille Toledano, PhD
MRC Centre for Environment and Health
Department of Epidemiology and Biostatistics
Imperial College London
St Mary's Campus
Norfolk Place, London
London, W2 1PG
United Kingdom
Phone: 44 02075943298
Email: m.toledano@imperial.ac.uk

Abstract

Background: The use of digital technology such as mobile phones is ubiquitous in adolescents. However, excessive use may have adverse health effects, possibly partially mediated by disruptions to sleep.

Objective: This study aims to assess the social predictors of digital technology use and their cross-sectional association with BMI z scores and being overweight in a large sample of adolescents.

Methods: We used baseline data from a subset of a large adolescent cohort from 39 schools across Greater London who participated in the Study of Cognition, Adolescents and Mobile Phones (n=1473). Digital technology use included phone calls, internet use on mobile phones, and video gaming on any device. Multilevel regression was used to assess the associations between digital technology use and age-specific and sex-specific BMI z scores and being overweight (including obesity). Measurements were derived from height and weight, obtained by the Tanita BC-418 Body Composition Analyzer. We examined whether these associations were mediated by insufficient sleep.

Results: Generally, participants with lower socioeconomic status reported more use of digital technology. Controlling for socioeconomic status, internet use on mobile phones for more than 3 hours per day was associated with higher BMI z scores (adjusted β=.30, 95% CI 0.11-0.48) and greater odds of being overweight (adjusted odds ratio 1.60, 95% CI 1.09-2.34), compared with low use (≤30 minutes). Similar associations were found between video gaming and BMI z scores and being overweight. The BMI z score was more strongly related to weekday digital technology use (internet use on mobile phones and video gaming) than weekend use. Insufficient sleep partly mediated the associations between digital technology use and BMI z scores (proportion of mediation from 8.6% to 17.8%) by an indirect effect.

Conclusions: We found an association between digital technology use and BMI in adolescents, partly mediated by insufficient sleep, suggesting that the underlying mechanisms may be multifactorial. Further research with longitudinal data is essential to explore the direction of the relationships.

(J Med Internet Res 2021;23(7):e26485) doi:10.2196/26485

XSL•FO
RenderX

KEYWORDS

adolescent; digital technology; obesity; insufficient sleep; mediation analysis; mobile phone

Introduction

Background

With advances in information and communication technologies, mobile phone use is ubiquitous, especially in young people, with 44% of 8- to 11-year olds and 86% of 12- to 15-year olds in the United Kingdom owning a mobile phone [1]. Adolescents aged 12-15 years also spend an average of 12 hours per week on video gaming [1]. Digital technology offers a broad range of functions, facilitating access to the internet and social networks, instant and text message exchange, multimedia, and entertainment. Despite unambiguous advantages such as easy access to information and fast communication, overuse of digital technology is prevalent in children and adolescents because of its powerful potential for diversion and escape from daily life [2,3]. This is becoming a major public health concern with adverse effects on physical health, psychological well-being, and academic achievement in adolescents [4-6].

Despite the exponential increase in digital technology use, the digital divide because of socioeconomic disparity is persistent. However, the nature of the divide today is a gap in the type and intensity of digital technology use, rather than a gap in access to digital devices [7]. Specifically, although family income is not associated with mobile phone ownership in adolescents [8], those with higher family income and parental education are less likely to show problematic mobile phone use or internet addiction as their digital behaviors (eg, time spent on digital devices and digital safety) may be supervised by their parents to a greater extent [9-12]. Adolescents who are marginalized or from disadvantaged groups tend to use digital technology at a higher rate, possibly because they are likely to experience dissatisfaction with their offline lives and web spaces can facilitate the development of relationships with peers with similar life experiences [13,14]. They may also have fewer opportunities to carry out extracurricular activities than their affluent peers and thus spend more time on digital devices [15]. However, there is a paucity of large-scale studies with participants from diverse socioeconomic backgrounds and nuanced exposure measurements to decipher the digital divide regarding various digital technology use, such as phone calls, internet use, and video gaming.

Previous studies have shown that longer duration of phone calls, internet use on mobile phones, and video gaming are associated with a higher risk of overweight and obesity in adolescents [16-19]. However, these studies have small sample sizes [18], do not specify weekday and weekend use [16,17], and only rely on self-reported height and weight, which is not reliable, especially for adolescents whose growth is rapidly changing [16,17,19]. Such limitations preclude a rigorous investigation of the relationship between digital technology use and obesity as well as the underlying mechanism, which is still unclear. The duration of screen-based device use, such as television, mobile phone, and video game consoles, is associated with insufficient sleep duration and increased sleep disturbance in adolescents [20-22]. Possible mechanisms include sleep time displacement by digital technology use and melatonin suppression at bedtime by light exposure emitted from these devices [23]. Insufficient sleep is a well-documented risk factor for obesity in adolescents through the development of insulin resistance, sedentariness, and unhealthy diet in terms of higher consumption of energy-dense foods and sweetened beverages [24]. Therefore, exploration of the complex interrelationships between digital technology use, sleep, and obesity in adolescents is crucial for understanding and preventing obesity.

We have recently established the Study of Cognition, Adolescents and Mobile Phones (SCAMP), the largest adolescent cohort study to investigate the use of mobile phones and other wireless devices and its association with cognitive, behavioral, health, and educational outcomes. We previously observed in this cohort that adolescents who used their mobile phones at nighttime, particularly in darkness, had poorer sleep outcomes compared with nonusers [22].

Objective

In this study, we aim to examine the associations between digital technology use (eg, phone calls, internet use on mobile phones, and video gaming) and measured BMI outcomes. In addition, we investigated whether insufficient sleep mediated these associations. We hypothesized that increasing digital technology use is associated with higher BMI z scores and being overweight and that the association between digital technology use and BMI z scores is mediated by insufficient sleep.

Methods

Study Design and Participants

SCAMP is a prospective adolescent cohort study, with details reported elsewhere [25]. Baseline data were collected from 6616 year 7 pupils (aged 11-12 years), recruited between 2014 and 2016, from 39 secondary schools (26 state and 13 independent) across Greater London, UK. Of all year 7 pupils (N=7375) at these 39 schools, 111 (1.51%) parents or pupils chose to opt out. The remaining nonparticipation (n=648) was because of absence, nonassents by parents, withdrawals, technical issues, or miscellaneous reasons.

Participants in all SCAMP schools completed a computer-based assessment using Psytools software (Delosis Ltd) in examination mode. The assessment included a questionnaire on their digital technology behaviors (eg, smartphone use, social media engagement, and video gaming); a battery of cognitive tests; and health, well-being, and behavior scales. A subset of SCAMP participants (12 out of 39 schools, n=2270) also participated in SCAMP *Bio-Zone* to provide noninvasive biological samples (urine and saliva) and anthropometric measurements (height, weight, waist circumference, grip, and pinch strength) as well as perform a lung function test. The SCAMP *Bio-Zone* participants with measured height and weight (n=1473) were included into the present analysis for assessing the association between digital technology use and BMI. Our sample (n=1473)

XSL•FO
RenderX

was powered to detect a small effect (effect size=0.005) for a two-tailed linear regression with 80% power. Our sample size is much larger than the required sample size (n=462) to detect a small mediation effect (eg, β=.14) with 80% power [26].

Exposure—Digital Technology Use

Categorical responses were provided for questions on digital technology use, for weekdays and weekends separately. Participants who reported using or having used a mobile phone were further asked about the duration of phone calls and internet use (eg, surfing the internet, WhatsApp, Facebook, YouTube, and any other web-based apps) on their mobile phones if their phones were able to connect to the internet. We examined both phone calls and internet use on mobile phones, given the vast majority of mobile phone owners in our study (5119/5490, 93.24%) had a mobile phone that connected to the internet. Internet use on tablets or laptops was not included, as it was not specifically measured. Participants were also asked about the duration of playing video games on any device. We integrated weekday and weekend use to yield average daily use, and categorized internet use on mobile phones and video gaming as 0-30 minutes, 31-59 minutes, 1-2 hours, and 3 hours or more, to allow for a sufficient sample size in each category. Daily use for ≥3 hours was defined as high use [27]. As very few participants reported long duration of phone calls, we categorized responses as 0-5 minutes, 6-15 minutes, 16-59 minutes, and ≥1 hour average daily use.

Mediator—Insufficient Sleep on Weekdays and Weekends

Adolescents provided information on when they usually got into bed, how long it took them to fall asleep, and what time they usually woke up, separately for weekdays and weekends. Sleep duration was derived from these responses. A sleep duration of less than 9 hours per day was defined as insufficient sleep, based on the recommendation of the US National Sleep Foundation for school-aged children [28].

Outcome—BMI z Score and Overweight Status at About 12 Years

We measured height (m) and weight (kg) using Tanita BC-418 Body Composition Analyzer. We used age- (in days) and sex-specific BMI z scores based on the 2007 World Health Organization growth reference by interpolating the World Health Organization references on a daily scale [29]. Participants with implausible BMI z scores (SD ≥4 away from the mean) were excluded (n=1). We also defined adolescent overweight and obesity as a BMI for age and sex corresponding to an adult BMI ≥25 and 30 kg/m^2, respectively, using the International Obesity Task Force cutoffs [30]. Owing to the relatively small number of adolescents with obesity (79/1473, 5.36%), we combined obese and overweight into one category of *overweight* (total n=377).

Covariates

Demographic information including age, sex, ethnicity (combined into *White*, *Black*, *Asian*, *Mixed*, and *Other*), and parental education and occupation was captured in the SCAMP assessment. Parental education was categorized in a binary form as follows: at least one parent with higher education and no parent with higher education. We used the Office for National Statistics classification of parental occupation, categorizing it into 3 levels [31]. Each child was allocated the higher level of either parent. Dietary factors were also considered as potential confounders. The participants were asked whether they normally eat breakfast. Responses were categorized as *yes* or *no*.

Statistical Analysis

Ordinal logistic regression yielded adjusted odds ratios (ORs) and 95% CIs for each exposure variable (phone calls, internet use on mobile phones, and video gaming) in relation to demographic characteristics and school type (state or independent). Multilevel linear regression was used to assess the associations between each of the exposure variables and BMI z score, adjusted for potential confounders and school clustering effect. Multilevel logistic regression was used to assess the associations between each exposure variable and overweight.

Potential confounders included age, sex, ethnicity, parental education, parental occupation, and breakfast eating, selected based on directed acyclic graphs (DAGs) [32]. DAG is a simple and transparent way to identify and demonstrate causal relationships between variables via graphs. Confounders were defined as the mutual causes of exposure and outcome variables. Model 1 showed unadjusted associations between each of the exposure variables and BMI z score or overweight. Model 2 was adjusted for age, sex, and ethnicity. In addition, we adjusted for socioeconomic status (SES) indicators (parental education and parental occupation) and breakfast eating in model 3. Missing values for covariates were assigned to a separate *missing* category for each covariate but were not excluded from the analysis. We also examined whether these associations were modified by sex, ethnicity, or parental occupation based on the significance of the interaction terms. We also assessed the associations between weekday and weekend digital technology use and BMI z scores. The z test for equality of regression coefficients was performed to assess whether associations with weekday use and weekend use differed.

Causal mediation analysis was performed to examine the role of insufficient sleep as a potential mediator between digital technology use and BMI z scores. Mediation effects of insufficient sleep on weekdays and weekends were analyzed separately. We used VanderWeele formula to compute the direct effect (ie, the effect of digital technology use on BMI z score independent of insufficient sleep), indirect effect (ie, the effect of digital technology use on BMI z score via insufficient sleep), total effect, and the proportion mediated by insufficient sleep [33]. SEs for mediation analysis were estimated using a bootstrap procedure with 5000 bootstrap replications to obtain a 95% CI. All analyses were performed using STATA version IC/13.1 for Windows (Stata Corp).

Ethical Approval

The North West Haydock Research Ethics Committee approved the SCAMP study protocol and subsequent amendments (reference number 14/NW/0347). Head teachers of schools consented to participation in SCAMP. Parents and adolescents

XSL•FO
RenderX

were provided with written information about the study in advance and were given the opportunity to opt out of the research at any time. The study was conducted in accordance with the Declaration of Helsinki.

Results

Overview

The median age of our study sample was 12.06 (IQR 11.79-12.33) years. The sample was diverse in terms of ethnicity and SES (Table 1). Almost a fifth of the cohort (1146/6616, 17.32%) spent ≥3 hours per day on internet use and 11.08% (733/6616) on video gaming. A total of 6.71% (444/6616) spent ≥1 hour per day on phone calls. The correlation between measures (eg, phone calls) of digital technology use was low (Cohen <0.3). Of the original 6616 participants in SCAMP, 1473 (22.26%) had measured height and weight. The mean BMI z score in our sample was 0.43 (SD 1.21). We could not find a substantial difference between these participants and the entire cohort in terms of sociodemographic characteristics and digital technology use [25].

XSL•FO
RenderX

Table 1. Sociodemographic characteristics, digital technology use (combining weekday and weekend use), and BMI z score of the Study of Cognition, Adolescents, and Mobile Phones cohort.

Sociodemographic characteristics	Overall (N=6616)	Bio-zone participants with measured height and weight (n=1473)
Age (years), median (IQR)	12.06 (11.79-12.33)	12.18 (11.94-12.44)
Sex, n (%)		
Male	3147 (47.57)	747 (50.71)
Female	3469 (52.43)	726 (49.29)
Ethnicity, n (%)		
White	2820 (42.62)	643 (43.65)
Black	1016 (15.36)	189 (12.83)
Asian	1758 (26.57)	425 (28.85)
Mixed	740 (11.19)	160 (10.86)
Other	62 (0.94)	17 (1.15)
Missing or not interpretable	220 (3.33)	39 (2.65)
Parental higher education, n (%)		
At least one	3677 (55.58)	891 (60.49)
None	1200 (18.14)	265 (17.99)
Missing	1739 (26.28)	317 (21.52)
Parental occupation, n (%)		
Managerial and professional	3426 (51.78)	867 (58.86)
Intermediate	1480 (22.37)	300 (20.37)
Routine and manual	1040 (15.72)	201 (13.65)
Missing	670 (10.13)	105 (7.13)
School type, n (%)		
State	5141 (77.71)	1035 (70.26)
Independent	1475 (22.29)	438 (29.74)
Phone calls, n (%)		
0-5 min	2236 (33.8)	559 (37.95)
6-15 min	1943 (29.37)	405 (27.49)
16-59 min	867 (13.1)	190 (12.9)
>1 h	444 (6.71)	81 (5.49)
Missing	1126 (17.02)	238 (16.16)
Internet use on mobile phones, n (%)		
0-30 min	2016 (30.47)	459 (31.16)
31-59 min	858 (12.97)	183 (12.42)
1-2 h	1094 (16.54)	253 (17.18)
>3 h	1146 (17.32)	264 (17.92)
Missing[a]	1502 (22.7)	314 (21.32)
Video gaming on any device, n (%)		
0-30 min	3440 (51.99)	788 (53.49)
31-59 min	981 (14.83)	233 (15.82)
1-2 h	1252 (18.92)	286 (19.42)
>3 h	733 (11.08)	124 (8.42)

XSL•FO
RenderX

Sociodemographic characteristics	Overall (N=6616)	Bio-zone participants with measured height and weight (n=1473)
Missing	210 (3.17)	42 (2.85)
BMI z score, mean (SD)	N/A[b]	0.43 (1.21)

[a]Participants who did not own a mobile phone were recorded as missing data.

[b]N/A: not applicable.

Associations Between Digital Technology Use and Other Variables

Table 2 shows that after mutual adjustment of all sociodemographic factors, participants at older ages reported longer duration of phone calls and internet use on mobile phones. Girls reported longer duration of phone calls but shorter duration of video gaming compared with boys. Compared with participants of White ethnicity, participants of Black ethnicity reported longer call duration and internet use on mobile phones, but the duration of video gaming was similar between White and Black participants. Asian participants reported lower levels of digital technology use. Participants whose parental occupation was classified as *routine and manual* reported longer duration of phone calls and video gaming than participants with parents in *managerial and professional* occupation. Participants from state schools reported longer duration of digital technology use than those from independent schools.

Table 3 shows that participants who spent ≥1 hour per day on mobile phone calls had 0.42 (95% CI 0.13-0.71) higher BMI z scores than those reporting <5 minutes of use, after adjusting for age, sex, ethnicity, parental education, parental occupation, and breakfast eating (model 3). Tables 3 and 4 show that participants who used the internet on mobile phones for more than 3 hours per day had 0.30 (95% CI 0.11- 0.48) higher BMI z scores and 60% ([1.6-1]/1) higher odds of being overweight (OR 1.6, 95% CI 1.09-2.34) than those reporting <30 minutes of use, respectively. Video gaming for ≥3 hours per day was also associated with higher BMI z scores (β=.26, 95% CI 0.03-0.5) and greater odds of being overweight (OR 1.62, 95% CI 1.03-2.53). The associations between digital technology use and BMI z score and being overweight were not modified by sex, ethnicity, or parental occupation (*P* values for interaction ranged from .08 to .89).

Table 5 shows that internet use on mobile phones and video gaming at high level (ie, more than 3 hours per day) on weekdays were associated with higher BMI z score (β=.30, 95% CI 0-0.61; β=.35, 95% CI 0.07-0.63, respectively). However, associations were not evident between high weekend use and BMI z score. The association between video gaming on weekdays and BMI z score was stronger than that between weekend use and BMI z score (*P*=.009). Stronger association of weekday internet use on mobile phones than weekend use with BMI z score was also observed, although the difference was not significant (*P*=.07).

Table 6 shows that insufficient sleep (both on weekdays and weekends) partly mediated (proportion of mediation ranged from 8.6% to 17.8%) the association between high digital technology use and BMI z score by an indirect effect. High use of digital technology remained associated with BMI z score after adjusting for insufficient sleep, shown as the direct effect.

XSL•FO
RenderX

Table 2. Associations between sociodemographic variables and digital technology use (combining weekday and weekend use) in the entire cohort (N=6616)[a].

Characteristics	Phone calls[b], aOR[c] (95% CI)	Internet use on mobile phones[d], aOR (95% CI)	Video gaming on any device[e], aOR (95% CI)
Age			
Per year increase	1.16 (1.02-1.32)	1.67 (1.46-1.90)	1.00 (0.88-1.14)
Sex			
Male	1 (reference)	1 (reference)	1 (reference)
Female	1.51 (1.36-1.67)	0.97 (0.88-1.08)	0.17 (0.15-0.19)
Ethnicity			
White	1 (reference)	1 (reference)	1 (reference)
Black	1.70 (1.46-1.97)	1.64 (1.41-1.90)	1.10 (0.95-1.27)
Asian	0.65 (0.57-0.74)	0.77 (0.67-0.88)	0.63 (0.56-0.72)
Mixed	1.24 (1.05-1.46)	1.24 (1.05-1.47)	1.24 (1.06-1.46)
Other	1.64 (0.99-2.73)	1.08 (0.64-1.82)	0.89 (0.55-1.44)
Parental higher education			
Yes	1 (reference)	1 (reference)	1 (reference)
No	0.99 (0.86-1.14)	1.12 (0.97-1.30)	1.04 (0.90-1.19)
Parental occupation			
Managerial and professional	1 (reference)	1 (reference)	1 (reference)
Intermediate	1.09 (0.95-1.24)	1.13 (0.98-1.29)	1.19 (1.04-1.36)
Routine and manual	1.19 (1.02-1.39)	1.13 (0.97-1.33)	1.27 (1.09-1.47)
School type			
Independent	1 (reference)	1 (reference)	1 (reference)
State	1.88 (1.65-2.15)	2.7 (2.35-3.11)	2.78 (2.40-3.22)

[a]Associations adjusted for all other independent variables in the table.

[b]Categorized as 0 to 5 minutes, 6 to 15 minutes, 16 to 59 minutes, and >1 hour.

[c]aOR: adjusted odds ratio, indicating the odds of being in higher categories of each characteristic of digital technology use associated with the independent variables relative to the reference group.

[d]Categorized as 0 to 30 minutes, 31 to 59 minutes, 1 to 2 hours, and >3 hours.

[e]Categorized as 0 to 30 minutes, 31 to 59 minutes, 1 to 2 hours, and >3 hours.

XSL•FO
RenderX

Table 3. Associations between digital technology use (combining weekday and weekend use) and BMI z score using multilevel linear regression (n=1473)[a].

Digital technology use	Model 1[b], β (95% CI)	Model 2[c], β (95% CI)	Model 3[d], β (95% CI)
Phone calls			
0-5 min	0 (reference)	0 (reference)	0 (reference)
6-15 min	.05 (−0.11 to 0.20)	.06 (−0.10 to 0.21)	.06 (−0.10 to 0.21)
16-59 min	.04 (−0.16 to 0.24)	.05 (−0.15 to 0.25)	.06 (−0.14 to 0.26)
>1 h	.40 (0.12 to 0.69)	.42 (0.13 to 0.71)	.42 (0.13 to 0.71)
P value for trend	.03	.03	.03
Internet use on mobile phones			
0-30 min	0 (reference)	0 (reference)	0 (reference)
31-59 min	.17 (−0.04 to 0.37)	.14 (−0.07 to 0.34)	.15 (−0.06 to 0.35)
1-2 h	.10 (−0.09 to 0.28)	.07 (−0.11 to 0.26)	.08 (−0.11 to 0.26)
>3 h	.31 (0.13 to 0.50)	.30 (0.11 to 0.49)	.30 (0.11 to 0.48)
P value for trend	.003	.005	.006
Video gaming on any device			
0-30 min	0 (reference)	0 (reference)	0 (reference)
31-59 min	.27 (0.10 to 0.45)	.23 (0.05 to 0.41)	.22 (0.04 to 0.40)
1-2 h	.18 (0.02 to 0.35)	.10 (−0.07 to 0.28)	.10 (−0.07 to 0.28)
>3 h	.37 (0.14 to 0.60)	.27 (0.03 to 0.50)	.26 (0.03 to 0.50)
P value for trend	<.001	.02	.03

[a]Age- and sex-specific BMI z score, using 2007 World Health Organization growth references for 5-19 years.

[b]Crude model.

[c]Adjusted for age, sex, and ethnicity.

[d]Additionally adjusted for parental education, parental occupation, and breakfast eating.

XSL•FO
RenderX

Table 4. Associations between digital technology use (combining weekday and weekend use) and overweight status using multilevel logistic regression (n=1473)[a].

Digital technology use	Model 1[b], OR[c] (95% CI)	Model 2[d], OR (95% CI)	Model 3[e], OR (95% CI)
Phone calls			
0-5 min	1 (reference)	1 (reference)	1 (reference)
6-15 min	1.03 (0.76-1.40)	1.03 (0.75-1.40)	1.02 (0.74-1.39)
16-59 min	1.18 (0.80-1.73)	1.2 (0.81-1.77)	1.23 (0.83-1.83)
>1 h	1.46 (0.87-2.45)	1.43 (0.84-2.43)	1.45 (0.85-2.48)
P value for trend	.15	.17	.15
Internet use on mobile phones			
0-30 min	1 (reference)	1 (reference)	1 (reference)
31-59 min	1.34 (0.89-2.01)	1.33 (0.88-2.00)	1.35 (0.89-2.05)
1-2 h	1.22 (0.84-1.77)	1.22 (0.84-1.78)	1.26 (0.86-1.85)
>3 h	1.55 (1.08-2.24)	1.57 (1.08-2.28)	1.60 (1.09-2.34)
P value for trend	.03	.03	.02
Video gaming on any device			
0-30 min	1 (reference)	1 (reference)	1 (reference)
31-59 min	1.42 (1.01-1.98)	1.38 (0.98-1.96)	1.38 (0.98-1.96)
1-2 h	1.35 (0.98-1.86)	1.29 (0.91-1.82)	1.29 (0.91-1.83)
>3 h	1.68 (1.11-2.56)	1.60 (1.03-2.49)	1.62 (1.03-2.53)
P value for trend	.005	.03	.03

[a]Overweight (including obesity) was defined as a BMI for age and sex corresponding to an adult BMI of ≥25 kg/m^2.

[b]Crude model.

[c]OR: odds ratio.

[d]Adjusted for age, sex, and ethnicity.

[e]Additionally adjusted for parental education, parental occupation, and breakfast eating.

XSL•FO
RenderX

Table 5. Associations between weekday and weekend digital technology use and BMI z score (n=1473)[a].

Digital technology use	Weekday use, β (95% CI)	Weekend use, β (95% CI)
Phone calls		
0-5 min	0 (reference)	0 (reference)
6-15 min	−.09 (−.29 to .10)	.01 (−.18 to .19)
16-59 min	−.13 (−.39 to .13)	.28 (.05 to .51)
>1 h	.13 (−.30 to .56)	.25 (−.11 to .61)
P value for trend	.77	.04
Internet use on mobile phones		
0-30 min	0 (reference)	0 (reference)
31-59 min	.16 (−.09 to .41)	0 (−.22 to .23)
1-2 h	.05 (−.23 to .33)	−.08 (−.33 to .18)
>3 h	.3 (.00 to .61)	−.02 (−.31 to .27)
P value for trend	.05	.73
Video gaming on any device		
0-30 min	0 (reference)	0 (reference)
31-59 min	.29 (.08 to .49)	−.09 (−.28 to .10)
1-2 h	.17 (−.08 to .41)	−.10 (−.31 to .12)
>3 h	.35 (.07 to .63)	−.11 (−.35 to .14)
P value for trend	.01	.40

[a]Adjusted for age, sex, ethnicity, parental education, parental occupation, breakfast eating, and school clustering effect; weekday and weekend use also mutually adjusted.

XSL•FO
RenderX

Table 6. Associations between digital technology use (combing weekday and weekend use) and BMI z score mediated by insufficient sleep (n=1473)[a].

Mediator and exposure	Indirect effect, β (95% CI[b])	Direct effect, β (95% CI)	Total effect, β (95% CI)	Proportion mediated (%)
Insufficient sleep on weekdays				
Phone calls				
0-5 min	0 (reference)	0 (reference)	0 (reference)	N/A[c]
6-15 min	.02 (−.01 to .02)	.03 (−.13 to .19)	.05 (−.13 to .19)	N/A
16-59 min	.03 (.00 to .06)	.10 (−.11 to .32)	.13 (−.08 to .34)	N/A
>1 h	.05 (.03 to .12)	.48 (.19 to .76)	.53 (.27 to .83)	8.6
Internet use on mobile phones				
0-30	0 (reference)	0 (reference)	0 (reference)	N/A
31-59	.02 (.00 to .05)	.12 (−.09 to .33)	.14 (−.07 to .35)	N/A
1-2 h	.03 (.01 to .06)	.06 (−.13 to .25)	.09 (−.10 to .27)	N/A
>3 h	.05 (.01 to .10)	.28 (.08 to .48)	.33 (.14 to .53)	15.9
Video gaming on any device				
0-30 min	0 (reference)	0 (reference)	0 (reference)	N/A
31-59 min	.01 (−.01 to .03)	.21 (.04 to .39)	.22 (.04 to .40)	N/A
1-2 h	.02 (.01 to .05)	.12 (−.05 to .31)	.14 (−.03 to .33)	N/A
>3 h	.05 (.02 to .09)	.26 (−.01 to .53)	.3 (.04 to .57)	15
Insufficient sleep on weekends				
Phone calls				
0-5 min	0 (reference)	0 (reference)	0 (reference)	N/A
6-15 min	.01 (−.01 to .02)	.04 (−.12 to .20)	.05 (−.12 to .20)	N/A
16-59 min	.03 (.00 to .05)	.1 (−.10 to .32)	.13 (−.08 to .34)	N/A
>1 h	.05 (.04 to .13)	.48 (.19 to .76)	.53 (.28 to .83)	9.2
Internet use on mobile phones				
0-30 min	0 (reference)	0 (reference)	0 (reference)	N/A
31-59 min	.01 (−.01 to .02)	.12 (−.09 to .33)	.13 (−.09 to .32)	N/A
1-2 h	.03 (.00 to .06)	.06 (−.13 to .24)	.09 (−.11 to .27)	N/A
>3 h	.05 (.00 to .09)	.29 (.08 to .48)	.33 (.12 to .52)	14.3
Video gaming on any device				
0-30 min	0 (reference)	0 (reference)	0 (reference)	N/A
31-59 min	.01 (−.01 to .03)	.22 (.04 to .39)	.22 (.05 to .41)	N/A
1-2 h	.02 (.01 to .07)	.13 (−.05 to .31)	.15 (−.02 to .34)	N/A
>3 h	.05 (.03 to .12)	.25 (−.03 to .52)	.31 (.04 to .58)	17.8

[a]Adjusted for age, sex, ethnicity, parent education, parent occupation, breakfast eating, and school type.

[b]CI was obtained by 5000 bootstrap resamples.

[c]N/A: not applicable.

Discussion

Principal Findings

Our results showed that Black participants, those with lower SES in terms of parental occupation, and those from state schools use digital technology to a greater extent than White participants, those with higher SES, and those from independent schools, respectively. This study also found that high use of digital technology, including internet use on mobile phones and video gaming, was associated with higher BMI z scores and greater odds of being overweight in adolescents, using objective height and weight measurements. The associations were consistent across sex, ethnicity, and SES. The BMI z score was more strongly related to high weekday digital technology use than high weekend use. Insufficient sleep on weekdays and

XSL•FO
RenderX

weekends partly mediated the associations between high digital technology use and BMI z scores.

In line with previous research [9,10], this study found that adolescents whose parents had management and professional occupations used digital technology for a shorter duration. These parents may be more aware of the adverse health consequences of excessive digital engagement and potentially exert more control on and guidance to their children to avoid overuse. Moreover, they may be equipped with advanced digital skills to manage their children's technology activities. Parents with lower SES, who can be under greater economic and time pressures, likely rate the restriction of children's screen time and support of children's web-based activities as relatively low priority compared with other life pressures [34]. Compared with White adolescents, Black adolescents used mobile phones for a longer duration. Black adolescents may have a sense of being marginalized in a school with primarily White students (as in our sample) [35]. The use of mobile phones to connect with family and friends is convenient to broaden and intensify their social support networks. This argument might also explain the different association patterns between ethnicity and video gaming, as the role of video gaming in social support is less pronounced than mobile phone use in general. Participants in state schools used digital technology more than those in independent schools, which is consistent with another study on British adolescents [3]. Although the state schools in our study generally have more restrictions on digital technology use than independent schools (ie, students are strictly not allowed to use their mobile phones during school time), participants from state schools might use digital technology more during weekday evenings.

Screen time, such as television watching, mobile phone use, and video gaming, has been associated with adolescent obesity [16,17,36]. However, inferences from these studies are limited because of the bias of self-reported weight and height as well as a lack of differentiation between weekday and weekend use. To our knowledge, this is the first study to distinguish between weekday and weekend use and investigate a plausible mechanism (insufficient sleep) to help explain the association between digital technology use and adolescent obesity.

This study adds to the literature by showing that weekday use of digital technology might play a more important role than weekend use in individual differences in BMI z scores. Previous research has found that biomarkers related to obesity (eg, insulin and Homeostatic Model Assessment for Insulin Resistance) are more strongly associated with weekday digital technology use than weekend use in adolescents [37]. This is possibly because of the metabolic risks attributed to prolonged sitting bouts already required during class time on weekdays [38]. This study also adds to the literature by showing that insufficient sleep (both on weekdays and weekends) partly explained the associations between digital technology use and BMI z scores through an indirect effect. Plausible explanations include that excessive use of digital technology may displace sleep time, particularly on weekdays. This is because most schools restrict the use of mobile phones during school time, and participants need to wake up early to attend school; therefore, the main time during which participants can use their phones on weekdays is during the evenings and nights. Although direct sleep time displacement is less likely on weekend nights, sleep loss might also be because of psychological and physiological arousal by the content as well as melatonin suppression caused by screen light [21,23]. Notably, the association between digital technology use and BMI z score was still evident after adjusting for insufficient sleep (ie, direct effect), indicating that the underlying mechanisms may extend beyond sleep deprivation. High engagement in digital technology may also displace physical activity, resulting in obesity [39]. Mobile phone and internet addiction, and video game dependency, may act as a manifestation of psychosocial stress, as digital technology may provide a means for adolescents to be distracted from stressful experiences [40-42]. The stress-induced hypothalamic-pituitary-adrenal axis could stimulate appetite and disrupt metabolism and thus promote weight gain [43,44].

This study has several strengths and weaknesses. It is based on the largest prospective cohort of adolescents worldwide, with detailed data on mobile phone technologies. The cohort is representative of school-aged children across Greater London. Analyses used objective measurements for height and weight and did not rely on self-reported information. The study was able to distinguish between weekday and weekend use and investigate a plausible mechanism of insufficient sleep to aid understanding of the association between digital technology use and adolescent obesity. However, this study is subject to residual confounding because of the nature of the observational study design, although confounders were rigorously selected on a scientific basis (ie, based on the DAG). In addition, information on physical activity, other dietary factors (eg, snacking), and psychosocial stressors was not available; therefore, other potential mechanisms could not be considered. We analyzed cross-sectional baseline data of the SCAMP sample; thus, we were not able to detect the temporal sequence of digital technology use and obesity. Longitudinal studies have shown that greater screen time is associated with increased changes in BMI and a higher risk of obesity in adolescents [19,45]. However, we cannot exclude the possibility of obesity leading to sedentary behaviors, such as high engagement in digital technology [46]. Baseline data were collected a few years ago and may, therefore, not reflect the contemporary digital environment, which is rapidly changing. However, we expect our findings to be relevant as screen time has been consistently associated with adolescent obesity in the literature [16,19,47,48]. The age range of our participants was small; therefore, our findings may not be generalizable to older adolescents. Finally, information on mobile phone use, video gaming, and sleep was derived from self-reports, possibly resulting in social desirability bias and measurement error.

This study suggests several avenues for future research. We plan to analyze longitudinal SCAMP data to investigate the temporal sequence of digital technology use and obesity and compare the relationship between digital technology use and obesity at different ages to create age-specific policy recommendations regarding digital technology activities. In addition, more comprehensive measurements of other obesity-related lifestyle factors such as physical activity and psychosocial stressors are important to unravel potential

XSL•FO
RenderX

biological and behavioral mechanisms linking digital technology use and adolescent obesity. Objective measures of screen time and a secondary measure of sleep collected using a wearable fitness tracker (eg, Fitbit monitor), in addition to self-reported information, are needed to minimize self-report bias.

Our findings also have significant public health implications. The evidence of the potential impacts of digital technology overuse on adolescent obesity via insufficient sleep will allow the development of guidance for children and adolescents' healthy use of digital technology to prevent obesity. It may be possible to maintain the level of digital technology use but mitigate its effects on health by altering patterns of use. In addition, the recent COVID-19 pandemic led to school closure, lockdown, and social distancing, which may substantially increase screen time and decrease opportunities for social and physical activity during this period [49]. Schools, teachers, and parents should be vigilant of the secondary impacts of COVID-19 on risks of sleep disturbance and obesity and better advise their students and children on digital technology use.

Conclusions

This study shows the socioeconomic disparities of digital technology use among 11-to 12-year olds. High use of digital technology, that is, >3 hours per day, was associated with higher BMI z scores and greater odds of being overweight. In general, BMI z scores were more strongly related to high weekday digital technology use than high weekend use. The association between digital technology use and BMI z score was partly mediated by insufficient sleep, suggesting that the underlying mechanisms are multifactorial. Further research on longitudinal data is essential to explore the direction of relationships. Schools, parents, public health practitioners, policy makers, and adolescents themselves should be aware of the potentially adverse health effects of excessive digital technology engagement on sleep and obesity.

Acknowledgments

The SCAMP study is an independent research funded by the National Institute for Health Research (NIHR) Policy Research Program (PRP; Secondary School Cohort Study of Mobile Phone Use and Neurocognitive and Behavioural Outcomes/091/0212) via the Research Initiative on Health and Mobile Telecommunications—a partnership between public funders and the mobile phone industry. This study was supported by funds from the NIHR Health Protection Research Units in Environmental Exposures and Health and Chemical and Radiation Threats and Hazards (Health Protection Research Units -2012-10141), a partnership between Public Health England and Imperial College London, and the MRC Centre for Environment and Health (MR/L01341X/1). An extension to SCAMP was funded by the NIHR PRP. The funders of the study had no role in the design or conduct of the study or in the reporting of the SCAMP study results. The views expressed in this paper are those of the authors and not necessarily those of the NIHR, Public Health England, or Department of Health and Social Care. The authors would like to thank all the schools, parents, and pupils who participated in the SCAMP. In addition, they would like to thank Dr Danielle Ashworth, Rosemary Jenkins, Yasmin Bou Karim, Irene Chang, Margaret Douglass, Mark Ellis, Dr Daniela Fecht, Charlotte Fleming, Prof John Gulliver, Dr Nick Henriquez, Rosi Hirst, Stacey Jennings, Daphna Kesary, Dr Gemma Knowles, Dr Michael O Mireku, William Mueller, Aamirah Mussa, Julian Mutz, Dr Jonathan Pham, Dr Milagros Ruiz, Annalisa Sheehan, Dr Rachel B Smith, and Riitta Soininen from Imperial College London; Dr Benjamin Barratt, Dr Artemis Doutsi, Dr Rosamund Dove, and Dr Ian Mudway from King's College London; Elizabeth Booth from Birkbeck, University of London; Dr John Rogers from Delosis Ltd, London; Mikaël J A Maes from University College London; Dr Marloes Eeftens from Swiss Tropical and Public Health Institute; and Dr Jürg Fröhlich and Marco Zahner from Swiss Federal Institute of Technology in Zürich for their contributions to the project.

Authors' Contributions

CS conceived the analysis plan, cleaned and analyzed the data, interpreted the results, and wrote and revised the manuscript. ID, M Thomas, PE, MR, and M Toledano obtained funding for the study and revised the manuscript. M Toledano conceived the study, obtained funding, and inputted the data analysis and manuscript writing. All authors have approved the manuscript for publication.

Conflicts of Interest

Both M Toledano and PE received research funding in support of research on mobile phones and health. PE and M Toledano are principal investigators of the UK Cohort Study of Mobile Phone Use and Health, which was funded by the UK Department of Health and Social Care PRP (grant PR-ST-0713-00003) and was formerly jointly funded by industry and government, via the independent Mobile Telecommunications and Health Research Program.

References

1. Children and parents: media use and attitudes report. Ofcom. 2017. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/108182/children-parents-media-use-attitudes-2017.pdf [accessed 2020-12-15]
2. Mak K, Lai C, Watanabe H, Kim D, Bahar N, Ramos M, et al. Epidemiology of internet behaviors and addiction among adolescents in six Asian countries. Cyberpsychol Behav Soc Netw 2014 Nov;17(11):720-728. [doi: 10.1089/cyber.2014.0139] [Medline: 25405785]

XSL•FO
RenderX

3. Lopez-Fernandez O, Honrubia-Serrano L, Freixa-Blanxart M, Gibson W. Prevalence of problematic mobile phone use in British adolescents. Cyberpsychol Behav Soc Netw 2014 Feb;17(2):91-98. [doi: 10.1089/cyber.2012.0260] [Medline: 23981147]
4. Punamäki RL, Wallenius M, Nygård CH, Saarni L, Rimpelä A. Use of information and communication technology (ICT) and perceived health in adolescence: the role of sleeping habits and waking-time tiredness. J Adolesc 2007 Aug;30(4):569-585. [doi: 10.1016/j.adolescence.2006.07.004] [Medline: 16979753]
5. Lemola S, Perkinson-Gloor N, Brand S, Dewald-Kaufmann JF, Grob A. Adolescents' electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. J Youth Adolesc 2015 Feb;44(2):405-418. [doi: 10.1007/s10964-014-0176-x] [Medline: 25204836]
6. Seo DG, Park Y, Kim MK, Park J. Mobile phone dependency and its impacts on adolescents' social and academic behaviors. Comput Hum Behav 2016 Oct;63:282-292. [doi: 10.1016/j.chb.2016.05.026]
7. Jackson LA, Zhao Y, Kolenic A, Fitzgerald HE, Harold R, Von Eye A. Race, gender, and information technology use: the new digital divide. Cyberpsychol Behav 2008 Aug;11(4):437-442. [doi: 10.1089/cpb.2007.0157] [Medline: 18721092]
8. Söderqvist F, Hardell L, Carlberg M, Hansson Mild K. Ownership and use of wireless telephones: a population-based study of Swedish children aged 7-14 years. BMC Public Health 2007 Jun 11;7(1):105 [FREE Full text] [doi: 10.1186/1471-2458-7-105] [Medline: 17561999]
9. Lau JT, Gross DL, Wu AM, Cheng K, Lau MM. Incidence and predictive factors of internet addiction among Chinese secondary school students in Hong Kong: a longitudinal study. Soc Psychiatry Psychiatr Epidemiol 2017 Jun 17;52(6):657-667. [doi: 10.1007/s00127-017-1356-2] [Medline: 28417158]
10. Sánchez-Martínez M, Otero A. Factors associated with cell phone use in adolescents in the community of Madrid (Spain). Cyberpsychol Behav 2009 Apr;12(2):131-137. [doi: 10.1089/cpb.2008.0164] [Medline: 19072078]
11. Männikkö N, Ruotsalainen H, Miettunen J, Marttila-Tornio K, Kääriäinen M. Parental socioeconomic status, adolescents' screen time and sports participation through externalizing and internalizing characteristics. Heliyon 2020 Feb;6(2):e03415 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03415] [Medline: 32099928]
12. Wang X, Xing W. Exploring the influence of parental involvement and socioeconomic status on teen digital citizenship: a path modeling approach. J Edu Technol Soc 2018;21(1):186-199 [FREE Full text]
13. Lin Y, Liu Q. Perceived subjective social status and smartphone addiction tendency among Chinese adolescents: a sequential mediation model. Child Youth Serv Rev 2020 Sep;116:105222. [doi: 10.1016/j.childyouth.2020.105222]
14. Devito M, Walker A, Birnholtz J, Ringland K, Macapagal K, Kraus A, et al. Social technologies for digital wellbeing among marginalized communities. In: Proceedings of the Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing. 2019 Presented at: CSCW '19: Computer Supported Cooperative Work and Social Computing; November 9-13, 2019; Austin TX USA p. 449-454. [doi: 10.1145/3311957.3359442]
15. La Torre G, Masala D, De Vito E, Langiano E, Capelli G, Ricciardi W, PHysical ActivitySocio-Economic Status collaborative group. Extra-curricular physical activity and socioeconomic status in Italian adolescents. BMC Public Health 2006 Jan 31;6(1):22 [FREE Full text] [doi: 10.1186/1471-2458-6-22] [Medline: 16448561]
16. Kenney EL, Gortmaker SL. United States adolescents' television, computer, videogame, smartphone, and tablet use: associations with sugary drinks, sleep, physical activity, and obesity. J Pediatr 2017 Mar;182:144-149. [doi: 10.1016/j.jpeds.2016.11.015] [Medline: 27988020]
17. Lajunen H, Keski-Rahkonen A, Pulkkinen L, Rose RJ, Rissanen A, Kaprio J. Are computer and cell phone use associated with body mass index and overweight? A population study among twin adolescents. BMC Public Health 2007 Feb 26;7(1):24 [FREE Full text] [doi: 10.1186/1471-2458-7-24] [Medline: 17324280]
18. Arora T, Hussain S, Lam KH, Yao GL, Thomas GN, Taheri S. Exploring the complex pathways among specific types of technology, self-reported sleep duration and body mass index in UK adolescents. Int J Obes (Lond) 2013 Sep 8;37(9):1254-1260. [doi: 10.1038/ijo.2012.209] [Medline: 23295500]
19. Mitchell JA, Rodriguez D, Schmitz KH, Audrain-McGovern J. Greater screen time is associated with adolescent obesity: a longitudinal study of the BMI distribution from Ages 14 to 18. Obesity (Silver Spring) 2013 Mar;21(3):572-575 [FREE Full text] [doi: 10.1002/oby.20157] [Medline: 23592665]
20. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Public Health 2011 Jan 31;11(1):66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]
21. Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015 Jun;21:50-58 [FREE Full text] [doi: 10.1016/j.smrv.2014.07.007] [Medline: 25193149]
22. Mireku MO, Barker MM, Mutz J, Dumontheil I, Thomas MS, Röösli M, et al. Night-time screen-based media device use and adolescents' sleep and health-related quality of life. Environ Int 2019 Mar;124:66-78 [FREE Full text] [doi: 10.1016/j.envint.2018.11.069] [Medline: 30640131]
23. Chang A, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. Proc Natl Acad Sci U S A 2015 Jan 27;112(4):1232-1237 [FREE Full text] [doi: 10.1073/pnas.1418490112] [Medline: 25535358]

XSL•FO
RenderX

24. Felső R, Lohner S, Hollódy K, Erhardt ?, Molnár D. Relationship between sleep duration and childhood obesity: systematic review including the potential underlying mechanisms. Nutr Metab Cardiovasc Dis 2017 Sep;27(9):751-761. [doi: 10.1016/j.numecd.2017.07.008] [Medline: 28818457]
25. Toledano M, Mutz J, Röösli M, Thomas M, Dumontheil I, Elliott P. Cohort Profile: The Study of Cognition, Adolescents and Mobile Phones (SCAMP). Int J Epidemiol 2019 Feb 01;48(1):25-26l [FREE Full text] [doi: 10.1093/ije/dyy192] [Medline: 30325429]
26. Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. Psychol Sci 2007 Mar 04;18(3):233-239 [FREE Full text] [doi: 10.1111/j.1467-9280.2007.01882.x] [Medline: 17444920]
27. Lane A, Harrison M, Murphy N. Screen time increases risk of overweight and obesity in active and inactive 9-year-old Irish children: a cross sectional analysis. J Phys Act Health 2014 Jul;11(5):985-991. [doi: 10.1123/jpah.2012-0182] [Medline: 23799255]
28. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
29. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 2007 Sep 01;85(9):660-667 [FREE Full text] [doi: 10.2471/blt.07.043497] [Medline: 18026621]
30. Cole T, Bellizzi M, Flegal K, Dietz W. Establishing a standard definition for child overweight and obesity worldwide: international survey. Br Med J 2000 May 06;320(7244):1240-1243 [FREE Full text] [doi: 10.1136/bmj.320.7244.1240] [Medline: 10797032]
31. The National Statistics Socio-Economic Classification (NS-SEC). Office for National Statistics. 2010. URL: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 [accessed 2020-12-15]
32. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37-48. [doi: 10.1097/00001648-199901000-00008]
33. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013 Jun;18(2):137-150 [FREE Full text] [doi: 10.1037/a0031034] [Medline: 23379553]
34. Lampard AM, Jurkowski JM, Davison KK. The family context of low-income parents who restrict child screen time. Child Obes 2013 Oct;9(5):386-392 [FREE Full text] [doi: 10.1089/chi.2013.0043] [Medline: 24004326]
35. Junco R, Merson D, Salter DW. The effect of gender, ethnicity, and income on college students' use of communication technologies. Cyberpsychol Behav Soc Netw 2010 Dec;13(6):619-627. [doi: 10.1089/cyber.2009.0357] [Medline: 21142986]
36. Jackson LA, von Eye A, Fitzgerald HE, Witt EA, Zhao Y. Internet use, videogame playing and cell phone use as predictors of children's body mass index (BMI), body weight, academic performance, and social and overall self-esteem. Comput Hum Behav 2011 Jan;27(1):599-604. [doi: 10.1016/j.chb.2010.10.019]
37. Hardy LL, Denney-Wilson E, Thrift AP, Okely AD, Baur LA. Screen time and metabolic risk factors among adolescents. Arch Pediatr Adolesc Med 2010 Jul 05;164(7):643-649. [doi: 10.1001/archpediatrics.2010.88] [Medline: 20603465]
38. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care 2008 Apr 05;31(4):661-666. [doi: 10.2337/dc07-2046] [Medline: 18252901]
39. Rey-López JP, Vicente-Rodríguez G, Biosca M, Moreno LA. Sedentary behaviour and obesity development in children and adolescents. Nutr Metab Cardiovasc Dis 2008 Mar;18(3):242-251. [doi: 10.1016/j.numecd.2007.07.008] [Medline: 18083016]
40. Lam LT, Peng Z, Mai J, Jing J. Factors associated with internet addiction among adolescents. Cyberpsychol Behav 2009 Oct;12(5):551-555. [doi: 10.1089/cpb.2009.0036] [Medline: 19619039]
41. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
42. Rehbein F, Kleimann M, Mößle T. Prevalence and risk factors of video game dependency in adolescence: results of a German nationwide survey. Cyberpsychol Behav Soc Netw 2010 Jan 14:269-277. [doi: 10.1089/cpb.2009.0227]
43. Hamer M, Steptoe A. Cortisol responses to mental stress and incident hypertension in healthy men and women. J Clin Endocrinol Metab 2012 Jan;97(1):29-34. [doi: 10.1210/jc.2011-2132] [Medline: 22031509]
44. Gundersen C, Mahatmya D, Garasky S, Lohman B. Linking psychosocial stressors and childhood obesity. Obes Rev 2011 May;12(5):54-63. [doi: 10.1111/j.1467-789X.2010.00813.x] [Medline: 21054757]
45. Boone JE, Gordon-Larsen P, Adair LS, Popkin BM. Screen time and physical activity during adolescence: longitudinal effects on obesity in young adulthood. Int J Behav Nutr Phys Act 2007 Jun 08;4(1):26 [FREE Full text] [doi: 10.1186/1479-5868-4-26] [Medline: 17559668]
46. Ekelund U, Brage S, Besson H, Sharp S, Wareham N. Time spent being sedentary and weight gain in healthy adults: reverse or bidirectional causality? Am J Clin Nutr 2008 Sep;88(3):612-617. [doi: 10.1093/ajcn/88.3.612] [Medline: 18779275]

XSL•FO
RenderX

47. Nightingale CM, Rudnicka AR, Donin AS, Sattar N, Cook DG, Whincup PH, et al. Screen time is associated with adiposity and insulin resistance in children. Arch Dis Child 2017 Jul 13;102(7):612-616 [FREE Full text] [doi: 10.1136/archdischild-2016-312016] [Medline: 28288985]
48. Laurson KR, Lee JA, Gentile DA, Walsh DA, Eisenmann JC. Concurrent associations between physical activity, screen time, and sleep duration with childhood obesity. ISRN Obes 2014 Mar 09;2014:204540-204546 [FREE Full text] [doi: 10.1155/2014/204540] [Medline: 24734210]
49. Nagata JM, Magid HS, Gabriel K. Screen time for children and adolescents during the coronavirus disease 2019 pandemic. Obesity (Silver Spring) 2020 Sep 07;28(9):1582-1583 [FREE Full text] [doi: 10.1002/oby.22917] [Medline: 32463530]

Abbreviations

DAG: directed acyclic graph
NIHR: National Institute for Health Research
OR: odds ratio
PRP: Policy Research Program
SCAMP: Study of Cognition, Adolescents and Mobile Phones
SES: socioeconomic status

Edited by R Kukafka; submitted 16.12.20; peer-reviewed by J Nagata, A Jolliff, S Smith; comments to author 08.02.21; revised version received 03.04.21; accepted 06.05.21; published 05.07.21.

Please cite as:
Shen C, Dumontheil I, Thomas M, Röösli M, Elliott P, Toledano M
Digital Technology Use and BMI: Evidence From a Cross-sectional Analysis of an Adolescent Cohort Study
J Med Internet Res 2021;23(7):e26485
URL: https://www.jmir.org/2021/7/e26485
doi:10.2196/26485
PMID:35143408

©Chen Shen, Iroise Dumontheil, Michael Thomas, Martin Röösli, Paul Elliott, Mireille Toledano. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Health Information Seeking Among African American and Hispanic Men Living With Chronic Conditions: Cross-sectional Survey Study

Ledric D Sherman[1], MA, PhD; Kirby Goidel[2], MA, PhD; Caroline D Bergeron[3], MSc, CHES, DrPH; Matthew Lee Smith[4], MPH, CHES, PhD

[1]Department of Health & Kinesiology, Texas A&M University, College Station, TX, United States

[2]Public Policy Research Institute & Department of Political Science, Texas A&M University, College Station, TX, United States

[3]Public Health Agency of Canada, Ottawa, ON, Canada

[4]Department of Environmental and Occupational Health, Center for Population Health and Aging, Texas A&M University, College Station, TX, United States

Corresponding Author:
Ledric D Sherman, MA, PhD
Department of Health & Kinesiology
Texas A&M University
4243 TAMU
College Station, TX, 77843
United States
Phone: 1 979 845 1266
Email: lsherman@tamu.edu

Abstract

Background: Previous research has identified disparities in seeking and using web-based health information to inform health-related behaviors. Relatively few studies however have examined the correlations between web-based health information seeking and use based on race, gender, age, and the presence of chronic health conditions.

Objective: In this study, we identify factors associated with seeking and using web-based health information among a uniquely vulnerable and intersectional population—middle-aged and older (40 years and older) African American and Hispanic men living with one or more chronic conditions.

Methods: Survey responses were collected from a purposive sample of African American and Hispanic men using Qualtrics web-based survey management software. To qualify for inclusion in the study, respondents had to identify as African American or Hispanic men, report having at least one chronic condition, and be aged 40 years and older. A series of binary logistic regression models was created using backward elimination. Statistical significance was determined at $P<.05$ for all analyses.

Results: Web-based health information seeking among African American and Hispanic men is a function of education, the presence of multiple chronic conditions, frustration with health care providers, internet use, and the perceived reliability of web-based health information. The use of web-based health information to inform interactions with health care providers was more common among African American and Hispanic men, who rated their health as relatively good, perceived barriers to care, used technology regularly, and took more daily medications.

Conclusions: Understanding the factors that influence African American and Hispanic men seeking web-based health information may help improve the care and treatment of chronic conditions. African American and Hispanic men seek web-based health information as a substitute for routine care and to inform their discussions with health care providers.

(J Med Internet Res 2021;23(7):e26180) doi:10.2196/26180

KEYWORDS

minority men; online information seeking; chronic disease; communication with health care providers; mobile phone

XSL•FO
RenderX

Introduction

Background

Information seeking encompasses the act of accumulating information to gain clarity or affirm knowledge about a specific topic [1]. Well-informed patients maintain a sense of control over their illness and are better able to cope with uncertainties related to outcomes and treatments [2-5]. Correspondingly, knowledgeable patients engage with medical providers in planning their care, managing their treatments, and adapting more readily to therapeutic schedules [6,7]. Insufficient health information, in contrast, can have unfavorable health consequences [2,8].

Health-related information can be obtained from supportive social networks, health care providers, and the media, including the internet, television, radio, books, or magazines [9,10]. Web-based health resources provide an *optimal way to disseminate health information* because there is the "immediacy of information access, the accessibility at any time of the day or night, the potential continual updating of information and the wider range of information available" [11]. However, disparities exist in terms of who pursues or seeks web-based health information and how the information is used to inform subsequent interactions with health care providers [7,12,13]. Although men use the internet more often than women, they use it less frequently to seek health information [14]. Men are also less likely to seek routine medical care than women and therefore have fewer opportunities to discuss web-based health information with their health care providers [15]. Race and ethnicity are also associated with internet-based health information seeking [16,17]. Historically, health information seeking was less common among racial and ethnic minorities because of limited internet access and lower health literacy skills [12,13,18-20]. However, recent research has suggested that these differences may be dwindling, with African Americans relying more heavily on web-based health information for health care [21,22].

Other factors associated with web-based information include education [19], self-reported health status [23,24], time spent with medical providers and frustrations in communicating with these providers [25,26], internet use [23,27], and the perceived reliability of web-based health information [28]. Previous research does not provide enough clarity on how these factors might affect internet-based information-seeking behaviors of African American and Hispanic men with chronic conditions; however, there is reason to expect some differences. Some studies have examined web-based health information seeking by race [29,30] or by sex, specifically for men with chronic diseases [31], but did not focus specifically on African American and Hispanic men with chronic conditions. This population has been found to experience important barriers to disease self-management [32], have less access to health insurance and preventative care [33], have higher rates of preventable hospitalizations [34], and are more likely to die from their chronic conditions compared with non-Hispanic White men [35]. Seeking and using credible web-based health information may represent an important health-promoting activity.

In this paper, we seek to understand web-based health information seeking among African American and Hispanic men. Our motivation is both substantive and methodological. First, African American and Hispanic men are less likely to seek preventative care and treatment, which subsequently affects health outcomes. Understanding the factors that lead to seeking web-based health information may lead to better health outcomes. Second, African American and Hispanic men are hard-to-reach populations in survey research, meaning that they are often underrepresented in probability-based samples.

Objective

In this 2-phase study, we contribute to the existing literature by investigating web-based health information seeking and use among African American and Hispanic men aged 40 years and older with one or more chronic conditions. In phase 1, we identify factors associated with seeking web-based health information in the past year about (1) a specific disease or medical problem and (2) medical treatments and procedures. Then, in phase 2, we identify factors associated with discussing web-based health information with primary medical professionals only among those men who sought health information on the internet and had a routine physician visit within the past year.

Methods

Overview

Due to increased costs and declining response rates, scholars increasingly rely on web-based panels when studying hard-to-reach or intersectional populations [36]. African American and Hispanic men with chronic conditions and aged 40 years or older, for example, are relatively small segments of the overall population, making random selection via probability sampling costly and inefficient. Due to distrust of medical providers, African American and Hispanic men are often less responsive to requests to participate in health-related research [37].

With this in mind, the sample in this study was designed using Qualtrics (Systems, Applications, and Products in Data Processing Societas Europaea) web-based panels to identify African American and Hispanic men aged 40 years and older with at least one chronic health condition. We used the Checklist for Reporting Results of Internet E-Surveys for web-based surveys in our description of data collection [38]. Qualtrics web-based panels are opt-in research panels ideal for studies targeting hard-to-reach populations. Qualtrics panels provide access to previously identified research participants with known characteristics, and panel participants are recruited and compensated for their participation by Qualtrics. Potential participants were directed to the programmed survey where they were provided with a description of the study and information relating to informed consent. The tradeoff for cost effectiveness using Qualtrics is that the sample might not be representative of the target population.

The survey questionnaire was constructed by the authors who identified validated questions from previous research related to web-based health information seeking and other health-related

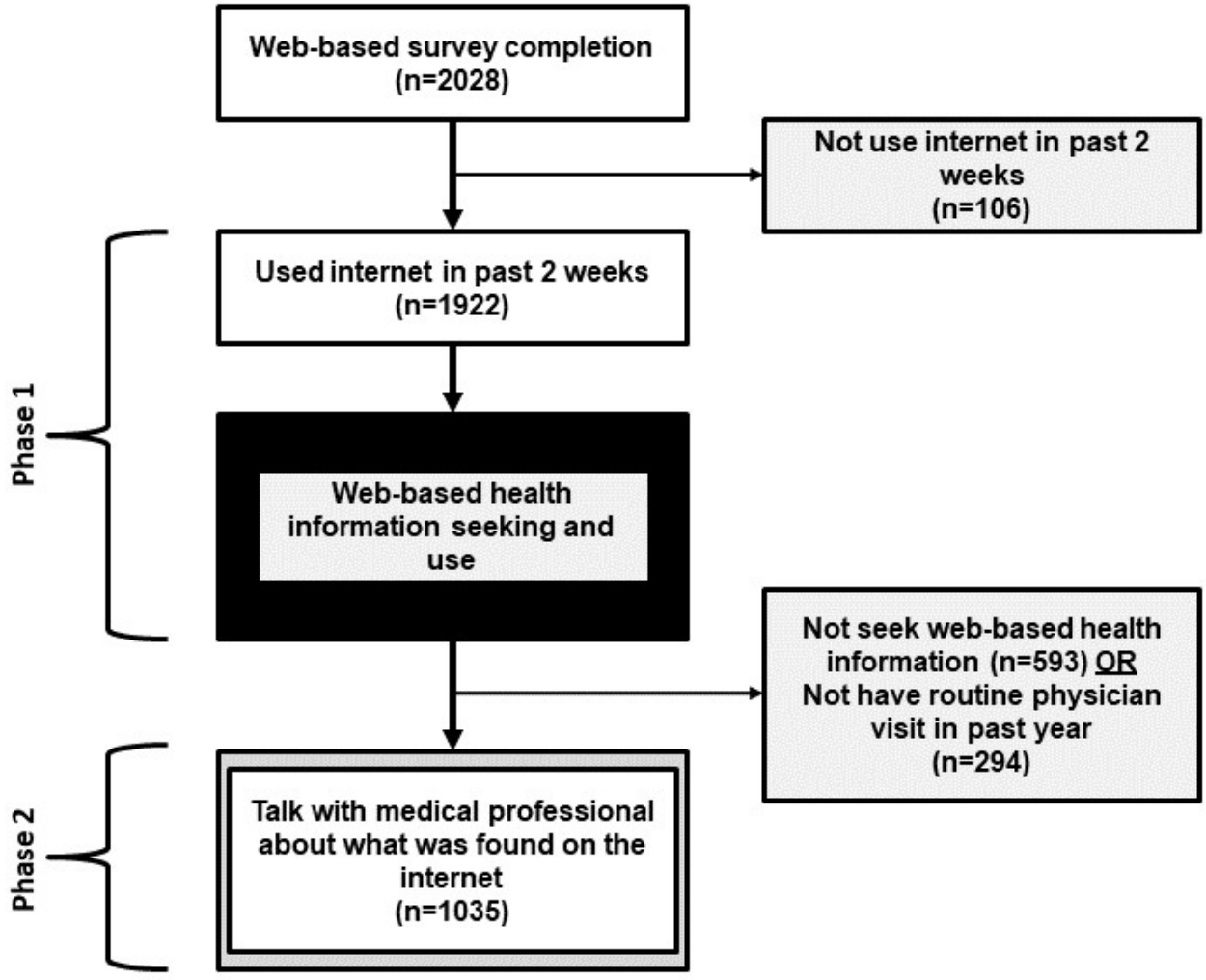
XSL•FO
RenderX

behaviors. An initial draft of the survey was carefully reviewed by experts in the field who were not a part of the research team and made suggestions for inclusion (or exclusion) of specific items. The final data were carefully reviewed by the research team, eliminating questionable responses (eg, respondents who completed the survey too quickly). In addition, filter questions for age, race, and the presence of one or more chronic conditions were used to further qualify potential respondents and ensure that only qualified respondents completed the survey questionnaire. The survey instrument included a wide range of health-related attitudes and behaviors. Overall, data were collected from 2028 men who met the inclusion criteria. This study was approved by the Institutional Review Board (#2018-1684) of Texas A&M University.

Inspired by Pettus et al [39], who examined internet use and web-based health information seeking among older women in a 2-phase study, we modeled this study to focus on web-based health information seeking among middle-aged and older men. To be included in phase 1 of our analytic sample, men also had to report using the internet within the past 2 weeks. Men who did not meet this criterion were excluded from phase 1 analyses. The 2 dependent variables in phase 1 assessed web-based health information seeking and use, which was measured using 2 items. First, participants were asked if they had looked for information on the internet about "a specific disease or medical problem." Response choices for this item were "yes" and "no." Second, participants were asked if they had looked for information on the internet about "a certain medical treatment or procedure." Response choices for this item were "yes" and "no."

Building upon findings in phase 1, to be included in phase 2 of our analytic sample, men must have reported "yes" to looking for information on the internet about "a specific disease or medical problem" or "a certain medical treatment or procedure. In addition, to avoid confounding the results with issues of health care access, these minority men must have reported having a routine physician visit in the past year to be included in phase 2. Men who did not meet this criterion were excluded from phase 2 analyses. The dependent variable in phase 2 assessed whether men shared findings from their web-based health information seeking with medical professionals. More specifically, participants were asked if they spoke with a medical professional about what they found on the web. Response choices for this item were "yes" and "no." Figure 1 illustrates the participant flow across both study phases based on the inclusion criteria.

Figure 1. Study flow by analysis phases.



XSL•FO
RenderX

We modeled web-based health information seeking as a function of demographics (age, race, education, marital status, and number of household members), health-related behaviors and status (number of chronic conditions, number of daily medications, having a routine physician visit in the past year, and self-reported health status), available resources for managing care (receiving help to manage care, ability to self-manage diseases, perceived barriers to care, health care frustrations, and participation in programs to prevent or manage chronic illness), and technology use and credibility (use of technology and reliability of web-based health information). We provide brief descriptions of each of these below.

Demographics

Age was measured in years, with all respondents reporting that they were aged 40 years or older. Race is a dichotomous variable indicating whether the respondent is African American (coded as 0 and serving as the baseline category) or Hispanic (coded as 1). Marital status was measured as a set of dummy variables indicating whether the respondent was single or never married, married or partnered, divorced or separated, or widowed. The number of household members was the total number of people (including the respondent) currently living in the household. The demographic variables in our models were primarily included as controls.

Health-Related Behaviors

This set of variables included health conditions, regular doctor visits, and self-reported health status. Overall, it is expected that individuals with worse health, meaning more chronic conditions and poor self-reported health status, would be more likely to look for health information on the internet. The number of chronic conditions was calculated using a "check all that apply" list of the following 19 chronic health conditions: (1) asthma, emphysema, chronic breathing problem, or lung problem; (2) arthritis or rheumatic disease; (3) cancer or cancer survivor; (4) chronic pain; (5) depression or anxiety; (6) diabetes; (7) heart disease; (8) high cholesterol; (9) hypertension; (10) kidney disease; (11) memory problem; (12) obesity; (13) osteoporosis; (14) obstructive sleep apnea; (15) schizophrenia or other psychotic disorder; (16) stroke; (17) thyroid problem; (18) urinary incontinence; and (19) another chronic condition not listed. In addition, participants were asked to report the number of different medications taken daily (range 0 to >6), whether they had visited a doctor in the past year (coded 1 if the respondent said yes; 0 otherwise), and a 5-point Likert scale measure of their self-reported health status ranging from poor (coded as 1) to excellent (coded as 0).

Resources for Managing Care

In addition to health-related behaviors and concerns, individuals with more resources available for managing care should be more likely to seek health information on the internet. This begins with perceptions of whether or not they are receiving the support they need to improve their health and manage their care, measured using a 5-point scale ranging from never (1) to always (5) [23,24]. Due to the skewed nature of the responses, these were collapsed into the never, rarely, or occasionally versus frequently or always range.

The disease self-management efficacy scale was included to gauge individual respondents' sense of control over the management of their health care [25,26], Respondents were asked about their level of agreement (using a 4-point Likert scale) with the following statements: (1) when all is said and done, I am the person who is responsible for taking care of my health; (2) taking an active role in my own health care is the most important thing that affects my health; (3) I know what each of my prescribed medications do; (4) I am confident that I can tell whether I need to go see the doctor or whether I can take care of a health problem myself; (5) I am confident I can tell a doctor concerns I have even if he or she does not ask; (6) I am confident I can follow through on medical treatments I may need to do at home; (7) I have been able to maintain (keep up with) lifestyle changes such as eating right or exercising; (8) I know how to prevent problems with my health; (9) I am confident I can figure out solutions when new problems arise with my health; and (10) I am confident that I can maintain lifestyle changes like eating right and exercising, even during times of stress. Scores for this scale ranged from 4 to 40, with higher scores indicating higher efficacy.

The use of web-based health information may also reflect barriers to care, reflecting the need for help and support in managing care and treatment. The barriers to self-care scale were measured based on levels of agreement with the following statements: (1) I need help learning what I should be doing to take better care of my health; (2) I need help learning how to take better care of my health in a way that works for me and my life; (3) I do not have the money it takes to do things that will improve my health or condition; (4) I wish I could change and do things that are healthier, but I just do not think I can; and (5) all of my different health problems and conditions make it difficult for me to take better care of myself. Scale values ranged from 5 to 20, with higher scores indicating more barriers.

Patients also seek web-based health information when their experiences with medical providers are frustrating. The health care frustration scale assesses whether participants felt any of the following frustrations [23,24]: (1) felt tired of describing their same conditions and problems every time they went to a hospital or doctor's office, (2) left the hospital or doctor's office and felt confused about what they should do, (3) wished their doctor had more time to spend talking with them, (4) felt tired of feeling on their own when it came to taking care of their health problems, (5) felt that their doctor did not realize what it was really like for them at home trying to take care of their health problems, and (6) wished they had a friend or family member who could go to the doctor with them. Responses were coded as "never" (1), "occasionally" (2), or "frequently" (3). Scores for this scale ranged from 6 to 18, with higher scores indicating higher health care frustrations [34].

Finally, respondents might gain knowledge about their chronic condition and insight into their medical condition by participating in a program specifically designed to prevent or treat chronic illnesses [40-43]. For example, the Chronic Disease Self-Management Program (CDSMP) is a universal program that applies to any chronic condition, although disease-specific translations also exist to build skills to manage arthritis, diabetes, chronic pain, and HIV and AIDS [44]. Previous research has

XSL•FO
RenderX

indicated that CDSMP improved outcomes while reducing costs [45].

Technology Use and Credibility

Aside from health concerns and conditions, web-based health information seeking is also a function of the level of comfort in using technology and perceptions regarding the credibility of information found alone. Technology use was measured by whether the respondent had used the following technologies in the past 2 weeks: computer (laptop, desktop, or tablet), smartphone, email (from a computer, smart phone, or tablet), internet (from a computer, smart phone, or tablet), Skype or other video systems (from a computer, smart phone, or tablet), or Facebook or other social media (eg, Twitter). Responses were coded from 0-6 depending on how many of these technologies individual participants reported having used in the past 2 weeks. Perceptions regarding the credibility of web-based information are measured with the question of how reliable they believe information on the internet is about health or medical conditions. Responses were coded from 0 (not at all) to 3 (extremely).

Data Analysis

All analyses were performed using SPSS version 25 (IBM Corporation). We calculated descriptive statistics for all variables of interest, which were compared across the 2 dependent variables in phase 1. Chi-square tests were used for categorical variables and two-tailed independent sample *t* tests were used for continuous and count variables, after assessing frequency distributions and tests for variance equality. As each dependent variable was dichotomous, we used logistic regression to estimate the models. Model selection was based on stepwise regression using backward elimination of nonsignificant predictor variables. Predictor variables were eliminated when they did not improve the overall model fit, as reflected by the likelihood ratio test. The final regression models included the fewest predictors from the model that provided the best fit to the data. Omnibus tests of model coefficients confirmed no significant loss of variance during backward entry steps for any of the 3 models fitted in this study. However, both full and final reduced regression models are presented in the tables described in the *Results* section. For all analyses, statistical significance was set at $P<.05$.

Results

Phase 1 Study Results

Table 1 provides the sample characteristics for the two phase 1 dependent variables. Among the 1922 men who had used the internet in the past week, 57.34% (1102) reported seeking information about a specific disease or medical problem and 50.83% (977) reported seeking information about a medical treatment or procedure. About 58.32% (1121/1922) of the participants were African American and 41.68% (801/1922) were Hispanic. The average age of the sample was 56.63 (SD 10.01) years. The majority of participants attended at least some college (1536/1922, 79.92%), over half were married or partnered (997/1922, 51.87%), and most reported having a routine physician visit in the past year (1627/1922, 84.65%). On average, participants reported living with 2.58 (SD 1.61) other people, having 3.93 (SD 2.9) chronic conditions, and taking 3.39 (SD 2.02) medications daily. About 57.7% (1109/1922) reported that they frequently or always received the help and support needed to improve their health and manage their health problems, and 17.43% (335/1922) reported attending a program to prevent or manage their chronic illness in the past year.

XSL•FO
RenderX

Table 1. Sample characteristics by web-based information-seeking behavior (N=1922).

Characteristics	Total (N=1922)	Looked for specific disease or medical problem					Looked for medical treatments and procedures				
		No (n=820)	Yes (n=1102)	Chi-square (*df*)	*t* test (*df*)	*P* value	No (n=945)	Yes (n=977)	Chi-square (*df*)	*t* test (*df*)	*P* value
Age (years), mean (SD)	56.63 (10.01)	57.77 (10.19)	55.78 (9.79)	N/A[a]	4.33 (1920)	<.001	57.22 (10.13)	56.05 (9.86)	N/A	2.56 (1920)	.01
Race or ethnicity, n (%)				3.1 (1)	N/A	.08			0.3 (1)	N/A	.59
African American	1121 (58.32)	497 (60.61)	624 (56.62)				557 (58.9)	564 (57.7)			
Hispanic	801 (41.68)	323 (39.39)	478 (43.38)				388 (41.05)	413 (42.27)			
Education, n (%)				20.2 (2)	N/A	<.001			19.3 (2)	N/A	<.001
High school or less	386 (20.08)	197 (24.02)	189 (17.15)				224 (23.7)	162 (16.58)			
Some college or 2-year degree	825 (42.92)	359 (43.78)	466 (42.29)				407 (43.06)	418 (42.78)			
4-year degree or more	711 (36.99)	264 (32.19)	447 (40.56)				314 (33.22)	397 (40.63)			
Marital status, n (%)				0.9 (3)	N/A	.82			1.5 (3)	N/A	.68
Married or partnered	997 (51.87)	427 (52.07)	570 (51.72)				481 (50.89)	516 (52.81)			
Never married	485 (25.23)	200 (24.39)	285 (25.86)				250 (26.45)	235 (24.05)			
Divorced or separated	365 (18.99)	162 (19.75)	203 (18.42)				178 (18.83)	187 (19.14)			
Widowed	75 (3.9)	31 (3.78)	44 (3.99)				36 (3.81)	39 (3.99)			
Persons living in household (including self), mean (SD)	2.58 (1.61)	2.47 (1.59)	2.67 (1.62)	N/A	−2.68 (1920)	.007	2.48 (1.57)	2.68 (1.63)	N/A	−2.76 (1920)	.006
Number of chronic conditions, mean (SD)	3.93 (2.9)	3.64 (2.76)	4.14 (2.98)	N/A	−3.8 (1920)	<.001	3.6 (2.73)	4.24 (3.02)	N/A	−4.85 (1920)	<.001
Number of medications taken daily, mean (SD)	3.39 (2.02)	3.36 (2.03)	3.42 (2.01)	N/A	−0.72 (1920)	.47	3.32 (2.05)	3.46 (1.98)	N/A	−1.51 (1920)	.13
Routine physician visit in past year, n (%)				0.2 (1)	N/A	.62			5.2 (1)	N/A	.02
No	295 (15.35)	122 (14.87)	173 (15.69)				163 (17.24)	132 (13.51)			
Yes	1627 (84.65)	689 (84.02)	929 (84.3)				782 (82.75)	845 (86.48)			
General health status, mean (SD)	2.85 (0.88)	2.91 (0.87)	2.81 (0.88)	N/A	2.55 (1920)	.01	2.89 (0.87)	2.82 (0.89)	N/A	1.78 (1920)	.08
Get the help or support needed to improve health and manage health problems, n (%)				8.3 (1)	N/A	.004			0.5 (1)	N/A	.48
Never or rarely or occasionally	813 (42.29)	316 (38.53)	497 (45.09)				392 (41.48)	421 (43.09)			
Frequently or always	1109 (57.7)	504 (61.46)	605 (54.9)				553 (58.51)	556 (56.91)			
Disease self-management efficacy (Cronbach α=.844), mean (SD)	28.54 (2.58)	28.58 (2.77)	28.5 (2.42)	N/A	0.69 (1920)	.49	28.61 (2.64)	28.47 (2.51)	N/A	1.22 (1920)	.22

XSL•FO
RenderX

Characteristics	Total (N=1922)	Looked for specific disease or medical problem					Looked for medical treatments and procedures				
		No (n=820)	Yes (n=1102)	Chi-square (*df*)	*t* test (*df*)	*P* value	No (n=945)	Yes (n=977)	Chi-square (*df*)	*t* test *(df)*	*P* value
Barriers to self-care (Cronbach α=.844), mean (SD)	11.48 (3.64)	10.96 (3.69)	11.87 (3.55)	N/A	−5.47 (1920)	<.001	11.04 (3.62)	11.91 (3.61)	N/A	−5.26 (1920)	<.001
Health care frustrations (Cronbach α=.856), mean (SD)	9.45 (3.13)	8.85 (2.93)	9.89 (3.2)	N/A	−7.4 (1920)	<.001	8.86 (2.9)	10.01 (3.24)	N/A	−8.2 (1920)	<.001
Sources of technology use in past 2 weeks, mean (SD)	4.98 (0.8)	4.87 (0.81)	5.07 (0.78)	N/A	−5.64 (1920)	<.001	4.85 (0.8)	5.12 (0.77)	N/A	−7.46 (1920)	<.001
Perceived reliability of information received on internet about health or medical conditions, mean (SD)	1.40 (0.69)	1.26 (0.69)	1.51 (0.68)	N/A	−7.75 (1920)	<.001	1.27 (0.68)	1.53 (0.68)	N/A	−8.4 (1920)	<.001
Ever attend program to prevent or manage chronic illness in past year, n (%)				22.4 (1)	N/A	<.001			38.7 (1)	N/A	<.001
No	1587 (82.57)	716 (87.31)	871 (79.04)				832 (88.04)	755 (77.27)			
Yes	335 (17.43)	104 (12.68)	231 (20.26)				113 (11.95)	222 (22.72)			

[a]N/A: not applicable.

When comparing sample characteristics by the 2 web-based health information–seeking behaviors (ie, both looked on the internet for information about specific diseases or medical problems and medical treatments and procedures), on average, participants who sought health information on the internet were significantly younger, lived with more people in their household, had more chronic conditions, reported more barriers to self-care, and reported higher health care frustrations. A significantly larger proportion of men who sought web-based health information were more educated and attended a program to prevent or manage their chronic illness in the past year. On average, participants who sought health information on the internet reported using more sources of technology and perceived health and medical information received on the internet to be more reliable.

Table 2 presents the results for seeking web-based information for a specific disease or medical condition among those reporting the use of the internet in the past 2 weeks. Compared with men who did not seek web-based health information for a specific disease or medical condition, men who had some college or a 2-year degree (odds ratio [OR] 1.35, 95% CI 1.04-1.74; *P*=.02), had a 4-year degree or higher (OR 1.91, 95% CI 1.45-2.50; *P*<.001), and attended a program to prevent or manage their chronic illness (OR 1.40, 95% CI 1.07-1.83; *P*=.01) were more likely to seek web-based information for a specific disease or medical condition. For each unit increase in self-reported chronic conditions (OR 1.04, 95% CI 1-1.08; *P*=.03), health care frustrations (OR 1.09, 95% CI 1.05-1.12; *P*<.001), sources of technology used (OR 1.27, 95% CI 1.12-1.44; *P*<.001), and perceived reliability of health and medical information received on the internet (OR 1.70, 95% CI 1.46-1.97; *P*<.001), the odds of seeking information on the internet for a specific disease or medical condition increased. For each unit increase in self-reported health status, the odds of seeking information on the internet for a specific disease or medical condition decreased (OR 0.86, 95% CI 0.76-0.97; *P*=.01).

XSL•FO
RenderX

Table 2. Factors associated with looking on the internet for information about a specific disease or medical problem (N=1922)[a].

Variable	Full model			Reduced model		
	β (SE)	*P* value	OR[b] (95% CI)	β (SE)	*P* value	OR (95% CI)
Age	−.01 (0.01)	.21	0.99 (0.98-1)	−.01 (0.01)	.09	0.99 (0.98-1)
Race or ethnicity						
African American	__[c]	—	1	—	—	1
Hispanic	.16 (0.10)	.12	1.17 (0.96-1.43)	.17 (0.10)	.09	1.19 (0.98-1.45)
Education						
High school or less	—	—	1	—	—	1
Some college or 2-year degree	.29 (0.13)	.03	1.33 (1.03-1.72)	.3 (0.13)	.02	1.35 (1.04-1.74)
4-year degree or more	.64 (0.14)	<.001	1.89 (1.43-2.49)	.65 (0.14)	<.001	1.91 (1.45-2.50)
Marital status						
Married or partnered	—	—	1	N/A[d]	N/A	N/A
Never married	−.02 (0.13)	.92	0.99 (0.76-1.28)	N/A	N/A	N/A
Divorced or separated	−.02 (0.14)	.89	0.98 (0.75-1.28)	N/A	N/A	N/A
Widowed	.20 (0.26)	.44	1.22 (0.74-2.03)	N/A	N/A	N/A
Persons living in household (including self)	.04 (0.03)	.25	1.04 (0.97-1.11)	N/A	N/A	N/A
Number of chronic conditions	.04 (0.02)	.049	1.04 (1-1.08)	.04 (0.02)	.03	1.04 (1-1.08)
Number of medications taken daily	.01 (0.03)	.63	1.01 (0.96-1.07)	N/A	N/A	N/A
Routine physician visit in past year						
No	—	—	1	N/A	N/A	N/A
Yes	.06 (0.15)	.69	1.06 (0.8-1.42)	N/A	N/A	N/A
General health status	−.13 (0.07)	.046	0.88 (0.77-1)	−.15 (0.06)	.01	0.86 (0.76-0.97)
Get the help or support needed						
Never or rarely or occasionally	—	—	1	N/A	N/A	N/A
Frequently or always	−.14 (0.11)	.22	0.87 (0.7-1.09)	N/A	N/A	N/A
Disease self-management efficacy	.01 (0.02)	.64	1.01 (0.97-1.05)	N/A	N/A	N/A
Barriers to self-care	.01 (0.02)	.38	1.01 (0.98-1.05)	N/A	N/A	N/A
Health care frustrations	.07 (0.02)	<.001	1.07 (1.03-1.12)	.08 (0.02)	<.001	1.09 (1.05-1.12)
Sources of technology use in past 2 weeks	.24 (0.06)	<.001	1.27 (1.12-1.44)	.24 (0.06)	<.001	1.27 (1.12-1.44)
Perceived reliability of information received on internet about health or medical conditions	.54 (0.08)	<.001	1.72 (1.48-1.99)	.53 (0.08)	<.001	1.7 (1.46-1.97)
Ever attend program to prevent or manage chronic illness in past year						
No	—	—	1	—	—	1
Yes	.31 (0.14)	.02	1.37 (1.04-1.79)	.34 (0.14)	.01	1.4 (1.07-1.83)

[a]Nagelkerke R^2=0.122 for full model; Nagelkerke R^2=0.119 (8 iterations) for reduced model.

[b]OR: odds ratio.

[c]Not available; referent category for independent variables.

[d]N/A: not applicable; referent category for dependent variable (not looking on the internet for information about a specific disease or medical problem).

Table 3 presents the results for seeking web-based information for medical treatments and procedures among those reporting having used the internet in the past 2 weeks. Compared with men who did not seek web-based information about medical treatments and procedures, men who had a college education or a 2-year degree (OR 1.32, 95% CI 1.02-1.72; *P*=.03), had a 4-year degree or higher (OR 1.72, 95% CI 1.31-2.25; *P*<.001), attended a routine physician visit in the past year (OR 1.48, 95% CI 1.13-1.94; *P*=.004), and attended a program to prevent or manage their chronic illness (OR 1.59, 95% CI 1.22-2.08; *P*=.001) were more likely to seek web-based information about medical treatments and procedures. For each unit increase in

XSL•FO
RenderX

self-reported chronic conditions (OR 1.06, 95% CI 1.02-1.1; *P*=.001), health care frustrations (OR 1.12, 95% CI 1.09-1.16; *P*=.001), sources of technology used (OR 1.44, 95% CI 1.27-1.63; *P*<.001), and the perceived reliability of health and medical information received on the internet (OR 1.69, 95% CI 1.46-1.95; *P*<.001), the odds of seeking web-based information about medical treatments and procedures increased.

Table 3. Factors associated with seeking on the internet information about medical treatments and procedures (N=1922)[a,b].

Variable	Full model			Reduced model		
	β (SE)	*P* value	OR[c] (95% CI)	β (SE)	*P* value	OR (95% CI)
Age	0 (0.01)	.77	1 (0.99-1.01)	N/A[d]	N/A	N/A
Race or ethnicity						
African American	__[e]	—	1	N/A	N/A	N/A
Hispanic	.05 (0.1)	.61	1.05 (0.86-1.29)	N/A	N/A	N/A
Education						
High school or less	—	—	1	—	—	1
Some college or 2-year degree	.28 (0.13)	.04	1.33 (1.02-1.72)	.28 (0.13)	.03	1.32 (1.02-1.72)
4-year degree or more	.54 (0.14)	<.001	1.72 (1.3-2.26)	.54 (0.14)	<.001	1.72 (1.31-2.25)
Marital status						
Married or partnered	—	—	1	N/A	N/A	N/A
Never married	−.21 (0.13)	.13	0.82 (0.63-1.06)	N/A	N/A	N/A
Divorced or separated	.05 (0.14)	.70	1.05 (0.81-1.38)	N/A	N/A	N/A
Widowed	.04 (0.26)	.89	1.04 (0.63-1.72)	N/A	N/A	N/A
Persons living in household (including self)	.04 (0.03)	.24	1.04 (0.97-1.11)	N/A	N/A	N/A
Number of chronic conditions	.05 (0.02)	.006	1.05 (1.02-1.1)	.06 (0.02)	.001	1.06 (1.02-1.1)
Number of medications taken daily	0 (0.03)	.99	1 (0.95-1.06)	N/A	N/A	N/A
Routine physician visit in past year						
No	—	—	1	—	—	1
Yes	.40 (0.15)	.007	1.49 (1.11-1.99)	.39 (0.14)	.004	1.48 (1.13-1.94)
General health status	−.07 (0.07)	.32	0.94 (0.83-1.07)	N/A	N/A	N/A
Get the help or support needed						
Never or rarely or occasionally	—	—	1	N/A	N/A	N/A
Get the help or support needed: frequently or always	.04 (0.11)	.70	1.04 (0.84-1.3)	N/A	N/A	N/A
Disease self-management efficacy	−.01 (0.02)	.52	0.99 (0.95-1.03)	N/A	N/A	N/A
Barriers to self-care	.02 (0.02)	.32	1.02 (0.98-1.05)	N/A	N/A	N/A
Health care frustrations	.1 (0.02)	<.001	1.11 (1.07-1.15)	.12 (0.02)	<.001	1.12 (1.09-1.16)
Sources of technology use in past 2 weeks	.37 (0.06)	<.001	1.45 (1.28-1.65)	.36 (0.06)	<.001	1.44 (1.27-1.63)
Perceived reliability of information received on internet about health or medical conditions	.56 (0.08)	<.001	1.75 (1.51-2.03)	.53 (0.07)	<.001	1.69 (1.46-1.95)
Ever attend program to prevent or manage chronic illness in past year						
No	—	—	1	—	—	1
Yes	.44 (0.14)	.001	1.56 (1.19-2.04)	.47 (0.14)	.001	1.59 (1.22-2.08)

[a]Nagelkerke R^2=0.155 for full model; Nagelkerke R^2=0.148 (10 iterations) for reduced model.

[b]The same dependent variable and referent category is used for the full and reduced models.

[c]OR: odds ratio.

[d]N/A: not applicable.

[e]Not available; referent category for independent variables.

XSL•FO
RenderX

Phase 2 Study Results

Table 4 presents phase 2 results for the 1035 participants who discussed what they found on the internet with their medical providers among those who had a routine physician visit in the past year. Relative to the 71.4% (739/1035) of men who reported both web-based health information seeking behaviors (ie, disease-specific information and medical treatments or procedures), men who only looked for information about specific diseases on the internet were significantly less likely to discuss what they found with their medical provider (OR 0.52, 95% CI 0.37-0.74; *P*<.001). Relative to men who did not discuss their web-based findings with medical providers, men who were Hispanic (OR 1.41, 95% CI 1.09-1.83; *P*<.001) and attended a program to prevent or manage their chronic illness (OR 2.19, 95% CI 1.61-2.98; *P*<.001) were more likely to discuss the web-based findings with their medical provider. For each unit increase in the number of medications taken daily (OR 1.13, 95% CI 1.05-1.21; *P*=.001), barriers to self-care (OR 1.04, 95% CI 1-1.08; *P*=.04), and sources of technology used (OR 1.24, 95% CI 1.05-1.46; *P*=.01), the odds of discussing web-based information with medical providers increased.

XSL•FO
RenderX

Table 4. Factors associated with discussing online information with medical providers (n=1035)[a,b].

Variable	Full model			Reduced model		
	β (SE)	*P* value	OR[c] (95% CI)	β (SE)	*P* value	OR (95% CI)
Looked on the internet for health information						
Both	__[d]	—	1	—	—	1
Only about medical treatments and procedures	−.35 (0.22)	.11	0.7 (0.45-1.09)	−.37 (0.22)	.09	0.69 (0.45-1.06)
Only about specific diseases of medical problems	−.6 (0.18)	.001	0.55 (0.39-0.79)	−.65 (0.18)	<.001	0.52 (0.37-0.74)
Age	0 (0.01)	.77	1 (0.99-1.02)	N/A[e]	N/A	N/A
Race or ethnicity						
African American	—	—	1	—	—	1
Hispanic	.36 (0.14)	.01	1.43 (1.09-1.87)	.35 (0.13)	.01	1.41 (1.09-1.83)
Education						
High school or less	—	—	1	N/A	N/A	N/A
Some college or 2-year degree	.23 (0.19)	.24	1.25 (0.86-1.83)	N/A	N/A	N/A
4-year degree or more	.17 (0.2)	.39	1.19 (0.8-1.76)	N/A	N/A	N/A
Marital status						
Married or partnered	—	—	1	N/A	N/A	N/A
Never married	−.11 (0.18)	.56	0.9 (0.63-1.28)	N/A	N/A	N/A
Divorced or separated	−.13 (0.19)	.51	0.88 (0.61-1.28)	N/A	N/A	N/A
Widowed	−.06 (0.34)	.86	0.94 (0.49-1.82)	N/A	N/A	N/A
Persons living in household (including self)	0 (0.05)	.99	1 (0.91-1.1)	N/A	N/A	N/A
Number of chronic conditions	.01 (0.02)	.73	1.01 (0.96-1.06)	N/A	N/A	N/A
Number of medications taken daily	.11 90.04)	.003	1.12 (1.04-1.21)	.12 (0.04)	.001	1.13 (1.05-1.21)
General health status	.1 (0.09)	.26	1.1 (0.93-1.31)	.15 (0.08)	.07	1.16 (0.99-1.35)
Get the help or support needed						
Never or rarely or occasionally	—	—	1	N/A	N/A	N/A
Frequently or always	.21 (0.15)	.16	1.23 (0.92-1.65)	N/A	N/A	N/A
Disease self-management efficacy	.02 (0.03)	.43	1.02 (0.97-1.08)	N/A	N/A	N/A
Barriers to self-care	.04 (0.02)	.1	1.04 (0.99-1.09)	.04 (0.02)	.04	1.04 (1-1.08)
Health care frustrations	.04 (0.03)	.11	1.04 (0.99-1.09)	N/A	N/A	N/A
Sources of technology use in past 2 weeks	.2 (0.09)	.02	1.22 (1.03-1.45)	.22 (0.09)	.01	1.24 (1.05-1.46)
Perceived reliability of information received on internet about health or medical conditions	.15 (0.1)	.15	1.16 (0.95-1.41)	N/A	N/A	N/A
Ever attend program to prevent or manage chronic illness in past year						
No	—	—	1	—	—	1
Yes	.7 (0.16)	<.001	2.01 (1.46-2.77)	.78 (0.16)	<.001	2.19 (1.61-2.98)

[a]Nagelkerke R^2=0.112 for full model; Nagelkerke R^2=0.101 (10 iterations) for reduced model.

[b]The same dependent variable and referent category is used for the full and reduced models.

[c]OR: odds ratio.

[d]Not available; referent category for independent variables.

[e]N/A: not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

This study examined health information seeking among a uniquely vulnerable and intersectional population, African American and Hispanic men aged 40 years and older with one or more chronic conditions. The specific results are worth discussing. First, internet-based health information is an important tool for African American and Hispanic men to use to learn about a specific disease or medical problem as well as medical treatments and procedures and to foster patient-provider conversations about these health-related internet searches, as illustrated by about half of the sample looking for information on the internet. Similar to previous studies, our study suggests that men who are younger [46-48], more highly educated [47,48], use technology more often [49], and believe the internet to be a reliable source [50] report seeking web-based health information in the past year to learn about a specific disease or medical treatment.

Those with more chronic conditions and greater health care frustrations were more likely to use the internet for both purposes (ie, to learn about a specific disease and medical treatments). Previous studies have demonstrated that people living with chronic conditions rely on the internet for help and support and might seek to learn about other people's experiences about a disease through web-based discussions [39]. People living with chronic conditions who experienced health care–related frustrations from unfulfilled needs in a medical encounter have been known to report greater functional limitations and greater self-care barriers to manage their condition or disease [51]. Internet-based health information could be used to help meet those needs.

Men who attended a disease prevention or management program in the past year were more likely to look on the internet to learn about a specific disease or medical problem and to learn about medical treatments and procedures. Considering that these evidence-based programs help increase participants' health behaviors and self-efficacy [52], program participants may feel encouraged to seek additional information to better understand the content of their program and their disease. Both web-based and traditional CDSMPs may also provide links, videos, and other resources to supplement the course materials, thereby encouraging program participants to seek this information on the internet as part of their disease self-management.

Interestingly, compared with men who used the internet only to learn about a specific disease or condition, men who reported both web-based health information–seeking behaviors were more likely to speak to their medical professional about what they found on the internet. It is possible that those who search on the web only to learn about a specific disease no longer feel the need to consult a physician, consequently substituting routine care [53]. Men who seek information on both diseases and treatments may have greater concerns about the credibility of web-based information or about their ability to evaluate this information [53]. Those who searched for treatments and procedures in addition to the general condition may also be exposed to web-based medical advertisements that encourage them to speak to their doctor about these treatment options [53-56]. Through their internet searches, patients reported increased confidence, control, and comfort in discussing their condition and treatments with their medical provider [57]; enhanced understanding of the medical jargon [58]; and satisfaction of feeling better informed [54]. Hispanic men more frequently discussed what they found on the internet with medical professionals. Studies suggest that web-based health information seeking gives Hispanic patients the confidence to discuss their health concerns with their doctors [30]. In the recent study by Camacho-Rivera et al [59] with a large representative sample of Hispanic adults in the United States, the authors found that Hispanics trusted cancer information from their doctors a lot (1014/1512, 67.06%) compared with information from the internet (309/1512, 20.44%). Although there was an important increase in trusting cancer information on the internet from 2014 to 2018, doctors remained the most trusted source of health information for Hispanics [59]. This study supports our findings that Hispanic men were more likely to talk with medical professionals about their web-based health searches.

Men with chronic conditions who had better general health statuses reported communicating with their medical professionals about what they found on the internet. This result contradicts previous studies [55], which suggest that those in poor health are more likely to talk to their medical providers about their web-based health information seeking than those in good health. However, higher medication intake is associated with poorer health (eg, frailty, disability, and fall risk) [60,61]. Poor health status can also lead to greater self-care barriers [62]. Those who take more medications daily also report more barriers to managing their chronic conditions [50]. Medication and self-care barriers were highlighted in our study as factors associated with discussing web-based health information with a medical professional.

It is possible that those with better health status, those who take more medications daily, and those with more self-care barriers may seek medical care more than once a year. It is known that increased physician visits to stay healthy and to get help to manage chronic conditions [63] may provide greater opportunities to discuss web-based health information–seeking behaviors. Increased visits may lead to better patient-physician interactions where bringing internet-based health information would not be seen as a threat but rather as something to be encouraged [64-66].

Limitations

This study has several limitations. The cross-sectional nature of this study did not allow for the assessment of causal relationships over time. On the basis of the funding mechanism supporting this study, data were only collected from African American and Hispanic men aged 40 years and older with one or more chronic conditions. Although these subgroups often report health-related disparities, additional insights might have been gained, including men of other races and ethnicities (eg, non-Hispanic White, Asian or Pacific Islander, and American Indian or Alaska Native). We hope that future research will expand the scope of this study and provide additional

XSL•FO
RenderX

comparisons. This study excludes African American and Hispanic males with one or more chronic conditions who do not have access to the internet. This digital divide continues to disproportionately impact the health of minorities and contribute to social inequalities in the United States [67]. In addition, no information was gathered about health literacy, the types of web-based information sources they used, or the credibility of these information sources (eg, government websites). Future research on internet-based health information–seeking behaviors of African American and Hispanic men with chronic conditions should consider assessing the health literacy level of respondents as well as their knowledge of credible health information sources [68]. In addition, in this study, African American and Hispanic subgroups were included in the analyses. Given the potential differences across racial or ethnic subgroups in terms of sociodemographics, behaviors, perceptions, and health care use, future studies may consider performing analyses on these subgroups separately or making direct comparisons between them.

Conclusions

Overall, this study provides an overview of health information–seeking behaviors among African American and Hispanic men with chronic conditions. Understanding these factors is crucial to influencing internet-based health communication, improving patient-provider communication, and ultimately improving the care and treatment of African American and Hispanic men.

Conflicts of Interest

None declared.

References

1. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
2. Evans SH, Clarke P. Flaws in health communication. Ann Int Commun Asso 2016 May 18;7(1):225-248. [doi: 10.1080/23808985.1983.11678537]
3. Felton BJ, Revenson TA. Coping with chronic illness: a study of illness controllability and the influence of coping strategies on psychological adjustment. J Consult Clin Psychol 1984;52(3):343-353. [doi: 10.1037/0022-006x.52.3.343]
4. Lerman C, Daly M, Walsh WP, Resch N, Seay J, Barsevick A, et al. Communication between patients with breast cancer and health care providers. Determinants and implications. Cancer 1993 Nov 01;72(9):2612-2620. [doi: 10.1002/1097-0142(19931101)72:9<2612::aid-cncr2820720916>3.0.co;2-f] [Medline: 8402483]
5. Viney LL, Westbrook MT. Coping with chronic illness: strategy preferences, changes in preferences and associated emotional reactions. J Chronic Dis 1984 Jan;37(6):489-502. [doi: 10.1016/0021-9681(84)90032-8]
6. Eraker SA, Kirscht JP, Becker MH. Understanding and improving patient compliance. Ann Intern Med 1984 Feb 01;100(2):258-268. [doi: 10.7326/0003-4819-100-2-258] [Medline: 6362512]
7. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
8. Denberg TD, Wong S, Beattie A. Women's misconceptions about cancer screening: implications for informed decision-making. Patient Educ Couns 2005 Jun;57(3):280-285. [doi: 10.1016/j.pec.2004.07.015] [Medline: 15893209]
9. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
10. Williams SL, Ames K, Lawson C. Preferences and trust in traditional and non-traditional sources of health information – a study of middle to older aged Australian adults. J Healthc Commun 2019 Jul 22;12(2):134-142. [doi: 10.1080/17538068.2019.1642050]
11. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
12. Ramanadhan S, Viswanath K. Health and the information nonseeker: a profile. Health Commun 2006 Jul;20(2):131-139. [doi: 10.1207/s15327027hc2002_4] [Medline: 16965250]
13. Tu H, Cohen GR. Striking jump in consumers seeking health care information. Track Rep 2008 Aug(20):1-8. [Medline: 18770913]
14. Hallyburton A, Evarts LA. Gender and online health information seeking: a five survey meta-analysis. J Cons Health Internet 2014 Jun 04;18(2):128-142. [doi: 10.1080/15398285.2014.902268]
15. Pinkhasov R, Wong J, Kashanian J, Lee M, Samadi DB, Pinkhasov MM, et al. Are men shortchanged on health? Perspective on health care utilization and health risk behavior in men and women in the United States. Int J Clin Pract 2010 Mar;64(4):475-487. [doi: 10.1111/j.1742-1241.2009.02290.x] [Medline: 20456194]
16. Rutten LJ, Squiers L, Hesse B. Cancer-related information seeking: hints from the 2003 Health Information National Trends Survey (HINTS). J Health Commun 2006 Feb;11 Suppl 1(sup001):147-156. [doi: 10.1080/10810730600637574] [Medline: 16641080]

XSL•FO
RenderX

17. Wiltshire J, Cronin K, Sarto GE, Brown R. Self-advocacy during the medical encounter: use of health information and racial/ethnic differences. Med Care 2006 Feb;44(2):100-109. [doi: 10.1097/01.mlr.0000196975.52557.b7] [Medline: 16434908]
18. Tardy RW, Hale CL. Getting "plugged in": a network analysis of health - information seeking among "stay - at - home moms". Commun Monogr 1998 Dec;65(4):336-357. [doi: 10.1080/03637759809376457]
19. Gidwani R, Zulman D. Association between acute medical exacerbations and consuming or producing web-based health information: analysis from Pew survey data. J Med Internet Res 2015 Jun 23;17(6):e145 [FREE Full text] [doi: 10.2196/jmir.3801] [Medline: 26104000]
20. Ackerson LK, Viswanath K. The social context of interpersonal communication and health. J Health Commun 2009 May 20;14 Suppl 1(sup1):5-17. [doi: 10.1080/10810730902806836] [Medline: 19449264]
21. Rooks RN, Kapral CG, Mathis AL. Chronic conditions may be more important than race or ethnicity in relation to health information seeking and use. J Aging Health 2019 Apr 01;31(4):611-630. [doi: 10.1177/0898264317744643] [Medline: 29254425]
22. Rooks RN, Wiltshire JC, Elder K, BeLue R, Gary LC. Health information seeking and use outside of the medical encounter: is it associated with race and ethnicity? Soc Sci Med 2012 Jan;74(2):176-184. [doi: 10.1016/j.socscimed.2011.09.040] [Medline: 22154611]
23. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016 Dec 26;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
24. Lee YJ, Boden-Albala B, Larson E, Wilcox A, Bakken S. Online health information seeking behaviors of Hispanics in New York City: a community-based cross-sectional study. J Med Internet Res 2014 Jul 22;16(7):e176 [FREE Full text] [doi: 10.2196/jmir.3499] [Medline: 25092120]
25. Bell RA, Hu X, Orrange SE, Kravitz RL. Lingering questions and doubts: online information-seeking of support forum members following their medical visits. Patient Educ Couns 2011 Dec;85(3):525-528. [doi: 10.1016/j.pec.2011.01.015] [Medline: 21315538]
26. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan 02;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
27. Lin CA, Atkin DJ, Cappotto C, Davis C, Dean J, Eisenbaum J, et al. Ethnicity, digital divides and uses of the internet for health information. Comput Hum Behav 2015 Oct;51:216-223. [doi: 10.1016/j.chb.2015.04.054]
28. Ye Y. Correlates of consumer trust in online health information: findings from the health information national trends survey. J Health Commun 2011 Jan 30;16(1):34-49. [doi: 10.1080/10810730.2010.529491] [Medline: 21086209]
29. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of HINTS data. Am J Health Promot 2021 Jan 26;35(1):84-92. [doi: 10.1177/0890117120934609] [Medline: 32588638]
30. Peña-Purcell N. Hispanics' use of internet health information: an exploratory study. J Med Libr Assoc 2008 Apr;96(2):101-107 [FREE Full text] [doi: 10.3163/1536-5050.96.2.101] [Medline: 18379664]
31. Sherman LD, Patterson MS, Tomar A, Wigfall LT. Use of digital health information for health information seeking among men living with chronic disease: data from the health information national trends survey. Am J Mens Health 2020 Jan 23;14(1):A [FREE Full text] [doi: 10.1177/1557988320901377] [Medline: 31973642]
32. Rustveld L, Pavlik VN, Jibaja-Weiss ML, Kline KN, Gossey T, Volk RJ. Adherence to diabetes self-care behaviors in English- and Spanish-speaking Hispanic men. Patient Prefer Adherence 2009 Nov 03;3:123-130 [FREE Full text] [doi: 10.2147/ppa.s5383] [Medline: 19936154]
33. National Research Council. In: Bulatao RA, Anderson NB, editors. Understanding Racial and Ethnic Differences in Health in Late Life: A Research Agenda. Washington, DC: The National Academies Press; 2004.
34. Laditka JN, Laditka SB. Race, ethnicity and hospitalization for six chronic ambulatory care sensitive conditions in the USA. Ethn Health 2006 Aug;11(3):247-263. [doi: 10.1080/13557850600565640] [Medline: 16774877]
35. Diabetes and African Americans. Department of Health and Human Services Office of Minority Health. 2015. URL: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=18 [accessed 2021-06-03]
36. Brick JM, Williams D. Explaining rising nonresponse rates in cross-sectional surveys. Ann Am Acad Polit Soc Sci 2012 Nov 26;645(1):36-59. [doi: 10.1177/0002716212456834]
37. Green BL, Partridge EE, Fouad MN, Kohler C, Crayton EF, Alexander L. African-American attitudes regarding cancer clinical trials and research studies: results from focus group methodology. Ethn Dis 2000;10(1):76-86. [Medline: 10764133]
38. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Dec 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
39. Pettus AJ, Mendez-Luck CA, Bergeron CD, Ahn S, Towne SD, Ory MG, et al. Internet-based resources for disease self-care among middle-aged and older women with chronic conditions. J Womens Health (Larchmt) 2017 Mar;26(3):222-233. [doi: 10.1089/jwh.2016.5843] [Medline: 27779440]
40. Lorig K, Holman H, Sobel D, Laurent D, González V, Minor M. Overview of self-management. In: Living a Healthy Life With Chronic Conditions: Self-management of Heart Disease, Arthritis, Diabetes, Asthma, Bronchitis, Emphysema and Others. Boulder, CO, United States: Bull Publishing Company; 2012:1-390.

XSL•FO
RenderX

41. Lorig K, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
42. Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001 Nov;39(11):1217-1223. [doi: 10.1097/00005650-200111000-00008] [Medline: 11606875]
43. Lorig KR, Ritter PL, González VM. Hispanic chronic disease self-management: a randomized community-based outcome trial. Nurs Res 2003;52(6):361-369. [doi: 10.1097/00006199-200311000-00003] [Medline: 14639082]
44. Self-Management Resource Center (SMRC). 2020. URL: https://www.selfmanagementresource.com/programs/ [accessed 2020-11-30]
45. Ory MG, Ahn S, Jiang L, Lorig K, Ritter P, Laurent DD, et al. National study of chronic disease self-management: six-month outcome findings. J Aging Health 2013 Oct 12;25(7):1258-1274. [doi: 10.1177/0898264313502531] [Medline: 24029414]
46. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33 [FREE Full text] [doi: 10.2196/jmir.1753] [Medline: 21752784]
47. Chung J, Gassert CA, Kim HS. Online health information use by participants in selected senior centres in Korea: current status of internet access and health information use by Korean older adults. Int J Older People Nurs 2011 Dec;6(4):261-271. [doi: 10.1111/j.1748-3743.2010.00238.x] [Medline: 21631875]
48. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
49. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a representative German health survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
50. Smith ML, Bergeron CD, Riggle SD, Meng L, Towne SD, Ahn S, et al. Self-care difficulties and reliance on support among vulnerable middle-aged and older adults with chronic conditions: a cross-sectional study. Maturitas 2017 Oct;104:1-10. [doi: 10.1016/j.maturitas.2017.06.030] [Medline: 28923169]
51. Smith ML, Bergeron CD, Adler CH, Patel A, Ahn S, Towne SD, et al. Factors associated with healthcare-related frustrations among adults with chronic conditions. Patient Educ Couns 2017 Jun;100(6):1185-1193. [doi: 10.1016/j.pec.2016.12.033] [Medline: 28117194]
52. Smith ML, Bergeron CD, Ahn S, Towne SD, Mingo CA, Robinson KT, et al. Engaging the underrepresented sex: male participation in Chronic Disease Self-Management Education (CDSME) programs. Am J Mens Health 2018 Jul 22;12(4):935-943 [FREE Full text] [doi: 10.1177/1557988317750943] [Medline: 29355070]
53. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
54. Kivits J. Informed patients and the internet: a mediated context for consultations with health professionals. J Health Psychol 2006 Mar 01;11(2):269-282. [doi: 10.1177/1359105306061186] [Medline: 16464924]
55. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
56. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of internet health information. Health Educ Res 2008 Jun 01;23(3):512-521. [doi: 10.1093/her/cyl062] [Medline: 16880222]
57. Broom A. Virtually he@lthy: the impact of internet use on disease experience and the doctor-patient relationship. Qual Health Res 2005 Mar 01;15(3):325-345. [doi: 10.1177/1049732304272916] [Medline: 15761103]
58. Chiu Y. Probing, impelling, but not offending doctors: the role of the internet as an information source for patients' interactions with doctors. Qual Health Res 2011 Dec 28;21(12):1658-1666. [doi: 10.1177/1049732311417455] [Medline: 21799204]
59. Camacho-Rivera M, Gonzalez CJ, Morency JA, Blake KD, Calixte R. Heterogeneity in trust of cancer information among hispanic adults in the United States: an analysis of the health information national trends survey. Cancer Epidemiol Biomarkers Prev 2020 Jul 01;29(7):1348-1356. [doi: 10.1158/1055-9965.epi-19-1375]
60. Gnjidic D, Hilmer SN, Blyth FM, Naganathan V, Waite L, Seibel MJ, et al. Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. J Clin Epidemiol 2012 Sep;65(9):989-995. [doi: 10.1016/j.jclinepi.2012.02.018] [Medline: 22742913]
61. Gnjidic D, Hilmer SN, Blyth FM, Naganathan V, Cumming RG, Handelsman DJ, et al. High-risk prescribing and incidence of frailty among older community-dwelling men. Clin Pharmacol Ther 2012 Mar 01;91(3):521-528. [doi: 10.1038/clpt.2011.258] [Medline: 22297385]
62. Sina M, Graffy J, Simmons D. Associations between barriers to self-care and diabetes complications among patients with type 2 diabetes. Diabetes Res Clin Pract 2018 Jul;141:126-131. [doi: 10.1016/j.diabres.2018.04.031] [Medline: 29689320]
63. Iverson S, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]

XSL•FO
RenderX

64. Bylund CL, Gueguen JA, D'Agostino TA, Li Y, Sonet E. Doctor-patient communication about cancer-related internet information. J Psychosoc Oncol 2010 Feb 22;28(2):127-142 [FREE Full text] [doi: 10.1080/07347330903570495] [Medline: 20391071]
65. Rahmqvist M, Bara A. Patients retrieving additional information via the internet: a trend analysis in a Swedish population, 2000-05. Scand J Public Health 2007 Sep 05;35(5):533-539. [doi: 10.1080/14034940701280750] [Medline: 17852982]
66. Stevenson FA, Kerr C, Murray E, Nazareth I. Information from the internet and the doctor-patient relationship: the patient perspective--a qualitative study. BMC Fam Pract 2007 Aug 16;8(1):47 [FREE Full text] [doi: 10.1186/1471-2296-8-47] [Medline: 17705836]
67. Campos-Castillo C. Revisiting the first-level digital divide in the United States: gender and race/ethnicity patterns, 2007–2012. Soc Sci Comput Rev 2014 Aug 21;33(4):423-439. [doi: 10.1177/0894439314547617]
68. Klawitter E, Hargittai E. "“I Went Home to Google”: How Users Assess the Credibility of Online Health Information". In: Hale TM, Chou WY, Cotten SR, Khilnani A, editors. eHealth: Current Evidence, Promises, Perils and Future Directions (Studies in Media and Communications, Vol. 15). Bingley, United Kingdom: Emerald Publishing Limited; 2018:11-41.

Abbreviations

CDSMP: Chronic Disease Self-Management Program
OR: odds ratio

Edited by R Kukafka; submitted 01.12.20; peer-reviewed by R Haase, S Song; comments to author 31.01.21; revised version received 11.04.21; accepted 04.05.21; published 14.07.21.

Please cite as:
Sherman LD, Goidel K, Bergeron CD, Smith ML
Web-Based Health Information Seeking Among African American and Hispanic Men Living With Chronic Conditions: Cross-sectional Survey Study
J Med Internet Res 2021;23(7):e26180
URL: https://www.jmir.org/2021/7/e26180
doi:10.2196/26180
PMID:34259646

©Ledric D Sherman, Kirby Goidel, Caroline D Bergeron, Matthew Lee Smith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions Toward the Use of Digital Technology for Enhancing Family Planning Services: Focus Group Discussion With Beneficiaries and Key Informative Interview With Midwives

Hind Yousef[1], MSCP, CBT, IPT; Nihaya Al-Sheyab[2], RN, PhD; Mohannad Al Nsour[1], MSc, MD, PhD; Yousef Khader[3], BDS, MSc, MSPH, MHPE, ScD; Malika Al Kattan[4], PhD; Marco Bardus[4], MA, PhD; Mohammad Alyahya[5], PhD; Hana Taha[1], PhD; Mirwais Amiri[1], MPH, MD

[1]Global Health Development | Eastern Mediterranean Public Health Network, Amman, Jordan

[2]Department of Allied Medical Sciences, Faculty of Applied Medical Sciences, Jordan University of Science and Technology, Irbid, Jordan

[3]Department of Community Medicine, Public Health and Family Medicine, Faculty of Medicine, Jordan University of Science & Technology, Irbid, Jordan

[4]Department of Health Promotion & Community Health, American University of Beirut, Beirut, Lebanon

[5]Department of Health Management and Policy, Jordan University of Science and Technology, Irbid, Jordan

Corresponding Author:
Hind Yousef, MSCP, CBT, IPT
Global Health Development | Eastern Mediterranean Public Health Network
4 Abu Al Ataheya St. Apt 5, Sport City
Amman
Jordan
Phone: 962 790883656
Email: hyousif@globalhealthdev.org

Abstract

Background: Modern family planning (FP) methods allow married couples to discuss and determine the number of children and years of spacing between them. Despite many significant improvements in FP services in Jordan, there are still many issues related to the uptake of FP services for both host communities and Syrian refugees, due to limitations in the health care system based on public health facilities. Digital technologies can provide opportunities to address the challenges faced in the health system, thus offering the potential to improve both coverage and quality of FP services and practices.

Objective: The aim of this study was to explore the perceptions of Jordanian women, Syrian refugees, and midwives in Jordan toward the use of digital health technology to support and enhance access to FP services.

Methods: We employed a qualitative study based on semistructured, face-to face key informative interviews with 17 midwives (providers) and focus group discussions with 32 married women of reproductive age (clients). Both midwives and clients were recruited from 9 health centers in 2 major governorates in Jordan (Irbid and Mafraq), where 17 in-depth interviews were conducted with midwives and 4 focus groups were conducted with the women. Each focus group included 4 Syrian refugees and 4 Jordanian women. The transcribed narratives were analyzed using inductive thematic analysis.

Results: Three major themes were derived from the narratives analysis, which covered the pros of using digital technology, concerns about digital technology use, and the ideal app or website characteristics. Ten subthemes emerged from these 3 main themes. Overall, midwives and women (Syrian refugees and host communities) agreed that digital technology can be feasible, cost-effective, well accepted, and potentially beneficial in increasing woman's awareness and knowledge regarding the FP methods and their side effect. Furthermore, digital technology can assist in enabling women's empowerment, which will allow them to make better decisions regarding FP use. No harmful risks or consequences were perceived to be associated with using digital technology. However, several concerns regarding digital technology use were related to eHealth literacy and the accuracy of the information provided. Midwives were mainly concerned about the patients who would rely mostly on the technology and choose to avoid consulting a health care professional.

Conclusions: As perceived by midwives and women, incorporating digital technology in FP services can be feasible, cost-effective, well accepted, and potentially beneficial in increasing woman's awareness regarding the FP methods and their side effect. It may also empower the women to play an active role in the shared (with their husband and family) decision-making process. Therefore,

XSL•FO
RenderX

digital technologies are recommended to address the challenges faced in health system and to improve both the coverage and the quality of FP services and practices.

(J Med Internet Res 2021;23(7):e25947) doi:10.2196/25947

KEYWORDS

family planning; mobile apps; social media; digital technology; contraceptives

Introduction

Family Planning Services in Jordan

Family planning (FP) is the necessary manifestation of the basic human rights where married couples get to determine the number of children and years of spacing between them [1,2]. FP is widely used throughout the world, but the uptake of the individual methods varies enormously between different countries and is relatively low in low- and middle-income countries (LMICs). In a recent scoping review, the unmet need for FP was reported to range from 20% to 58% across 34 studies conducted in LMICs [3]. According to the World Health Organization (WHO), FP uptake is generally low in some specific segments of the population, such as refugees, women in the postpartum period, adolescents, migrants, and urban slum dwellers [4]. The United Nations High Commissioner for Refugees (UNHCR) conducted a multicountry assessment on FP services offered to refugees in Bangladesh, Djibouti, Jordan, Kenya, Malaysia, and Uganda [5]. The assessment showed that 15-19-year-old married refugee women had higher unmet need, were less likely to use contraception, and had lower awareness compared with older women aged 20-49 years.

In Jordan, while a 2016 study found that 42% of married Syrian refugees never used modern FP methods [6], the latest population and family health survey according to the Demographic and Health Surveys program showed that the unmet need for contraception among Syrian refugees was 19%, compared with 14% among Jordanian women [7]. Jordan's fertility rate showed only a minor decline from 3.7% in 2002 to 3.5% in 2012 and a further decline to 2.7% in 2017 [1]. This drop was significant considering the noticeably high birth rates among Syrian refugees, which could have indirectly increased the overall Jordanian fertility rates in the last decade [2,4]. The main factor attributed to the reduction in fertility rate in Jordan was the adoption of FP strategies among Jordanian families [8].

Accessing FP and health services is difficult in countries of the Eastern Mediterranean, such as Jordan and Lebanon, whose health care systems had to sustain the influx of large numbers of Syrian refugees. According to the UNHCR, as of June 2019, the number of registered Syrian refugees exceeds 664,000 in Jordan [9]. In this country, the Ministry of Health provides free primary health care, including maternal and child health, to all residents including Syrian refugees. Yet, the use of any FP method has fallen from 61% to 52% between 2012 and 2017 with more reduction in traditional FP methods (from 23% to 11%) [7]. Surprisingly, more than half of Jordanian women of childbearing age reported not using any type of modern FP methods [3], a fact that highlights the gap between the reality and the Sustainable Development Goals' intention of improving the use of modern FP methods to 80% [4]. Further, despite the significant improvement in FP services through providing free FP methods and counseling to both host communities and Syrian refugees, there are many issues that hamper the access to FP services due to the manner in which public health facilities operate. For instance, failure to adapt to the latest technology is one of those issues, as it has not been able to bridge the gender role gap between women and men culturally.

Barriers to Accessing Family Planning Services

According to a recent UNHCR report [5], barriers to FP use among refugees globally include availability, accessibility, and acceptability of FP services. Acceptability barriers include, for example, religious opposition, perceived social pressure from family members, and language barriers in dealing with local health care providers; availability of services and accessibility barriers include, for example, distance from health care facilities and costs of transport and of a doctor's consultation [5]. These factors include also discrimination and other biases from the providers, which impact the quality of care, or perceived social pressure related to large expected family sizes from spouses, parents, and relatives [10]. Among Jordanian and Syrian women, available research shows that common barriers to FP use include fear of adverse side effects and complications [11], which leads to uptake of less effective traditional methods (eg, withdrawal and condoms). This was coupled with limited awareness about fertility limitation and FP in general [12,13], and limited sexual health knowledge [14]. From a sociocultural perspective, most married couples, being both Jordanian and Syrian, believe that Islam promotes birth spacing while encouraging having a large family size. As a matter of fact, the Arab culture appears to be a significant determinant of family size, having preference for large families and male children, facts that can have a significant influence on FP decisions [15].

The Potential of Digital Technologies for Family Planning

Using digital technologies for health promotion has become a prominent practice for utilizing both routine and innovative information and communications technology in addressing health needs [16]. Various reviews focusing on sexual and reproductive health interventions showed that mobile technologies can improve the uptake of services [17,18]. For example, SMS text messages can be used as reminders to improve attendance to doctors' appointments and compliance with medications [19-21]. In LMICs, some studies have tested the use of mobile SMS text messages to reduce unwanted pregnancy among clients [22], or the usage of mobile-delivered health communication campaigns to encourage discussion about FP [23]. Digital health interventions were also used to provide FP counseling with a mobile job aid [24,25] and training for health care workers using SMS text messages and interactive

XSL•FO
RenderX

voice response [26]. However, limited evidence exists on the use of these technologies in the Arab world and among refugees [27,28] and for the purpose of promoting FP services.

According to the Global Digital Health Index (GDHI) [29], Jordan has a fully functional governance structure that monitors the implementation of digital health strategies. The workforce lacks some training in digital health and thus requires preservice training. Similarly, digital technologies can provide opportunities to address the challenges faced in the health system, thus offering the potential to improve both the coverage and quality of health services and practices [16]. Some research has shown that women are increasingly interested in contraceptive tools utilizing mobile technology, and most women expect them to be science based [30,31]. The majority of mobile apps available support natural FP methods, which are recognized as the least effective FP method [32]. However, the implementation of digital health technologies has been extensively used only in the short term and without rigorous examination of benefits and harms on the health system and people's quality of life [33]. Moreover, health care providers and stakeholders need to understand what motivates and hinders people to use digital health interventions [32]. While some qualitative studies have explored the perceptions toward FP among community members and health providers in LMICs or Global South countries such as South Africa [33], to the best of our knowledge, no previous studies have been conducted in Jordan to explore the perceptions of the midwives and clients (Jordanian women and Syrian refugees) toward digital technology use for promoting FP [31].

Objectives

This study aimed to explore the perceptions of Jordanian women, Syrian refugees, and midwives in Jordan toward the use of digital health technology to support access to FP services. This study constitutes the formative research activity to develop a digital solution to increase the uptake of FP services among Syrian refugees and host communities in Jordan.

Methods

Study Design

Overview

A qualitative research was conducted among 2 study populations: Jordanian women and Syrian refugees (the clients), and midwives who provide FP services (the providers). We conducted focus group discussions (FGDs) with clients and one-to-one, face-to-face in-depth interviews with the providers. The objective of the focus groups and interviews was to explore the participants' perceptions about using digital technology (such as mobile apps, SMS text messages, and other tools) as a means to receive counseling and obtain information on FP.

FGDs With Clients

Data were collected between January and March 2020. The research team recruited a purposive sample (n=32) of married women of reproductive age through the midwives working at the International Rescue Committee (IRC) and Comprehensive Health Centers in 2 governorates in northern Jordan. Four FGDs were conducted among women attending the IRC clinics and Ministry of Health primary care facilities. Each FGD included 4 Jordanian and 4 Syrian women. To obtain a broader view of the participants' perceptions about FP, women of different ages and different levels of education were invited. It is important to consider that it was initially planned to conduct 15 FGDs with clients. However, after conducting 4 FGDs, the research team realized saturation was reached and sufficient information was obtained.

The research team developed an FGD interview guide. A moderator guided the discussion and encouraged all women to share their perception and to express their own views. Comprehension probes were used if needed to clarify responses. The focus groups lasted from 90 to 120 minutes.

Interviews With Providers

We conducted in-depth interviews with all the midwives (n=17) who provide FP services in 9 different health centers in 2 major governorates in Jordan: Irbid and Mafraq. Two of the 9 health centers primarily provide services for Syrian women. All the midwives were previously trained to provide FP counseling and services. Data were collected during January-March 2020. Two trained female investigators (HY and NA-S) conducted the individual interviews with the midwives using an interview guide designed by the research team. This guide covered the major concepts to be discussed during the interviews. Other more specific questions and probes were also asked, as appropriate, but were not initially included in the interview guide but emerged during the active discussion. All interviews were held in a quiet setting at a convenient place to all midwives after obtaining their approval and consent. Researchers conducted the interviews in the local Arabic dialect. Interviews lasted for around 30-45 minutes depending on the midwives' level of interaction and sharing experiences.

A digital voice recorder was used for the FGDs and the face-to-face interviews, as it allowed easy management of interviews and it recorded high-quality audio, which facilitated the transcription.

Ethical Approval

The study received ethical approval from the Research Ethics Committee of the Jordan University of Science and Technology and the Ministry of Health in Jordan (Ref.: 6/127/2019, received on September 12, 2019). During participant recruitment and before data collection, study participants were given verbal and written information about the study aims and objectives. Furthermore, they were informed that their participation in this study was confidential and voluntary, and that they could withdraw from the study at any given point without negative consequences to them. All participants signed a consent form, which included the consent to audiotape the interviews and FGDs.

Data Analysis

All discussions were transcribed verbatim. Data were analyzed using inductive thematic analysis [34,35]. Preliminary analysis was conducted after each interview to get a general impression of the results, which allowed for early identification of areas

XSL•FO
RenderX

that needed additional clarifications from the study participants. Researchers conducted the thematic analysis in its original language (Arabic) to maintain trustworthiness and credibility of the findings, which could have been lost by initial, inaccurate translation [36]. Translation into English was commenced after themes were generated and for extraction of quotations. The initial phase of the thematic analysis, as described by Braun and Clarke [34] "becoming familiar with the data," was initiated by reading and re-reading the transcripts several times. Repeated reading contributed to get a better understanding and enhanced researchers' familiarity with the data. Following the initial stage, *codes were generated* from the data set. Coding was carried out by systematically organizing and gaining meaningful characteristics of data related to the research question. One of the researchers (HA) began the process of initial coding; all transcriptions were coded one by one. This process was repeated until coding consensus was reached. For the transcribed data from FGDs, data analysis was undertaken manually through coding and generating categories and themes.

Concerning the transcripts for the face-to-face interviews with midwives, Qualitative Data Analysis in R (RQDA) was used for analysis. The third phase "searching for themes" focused on a broader level of analysis and involved the researcher identifying suitable themes to which codes could be attributed [34,35]. Initial codes pertinent to research question were integrated into themes considering how relationships were formed between codes and potential themes. To visualize and explore trends and relationships in the source data, codes, and themes, a tree mapping was formulated using RQDA. Derived themes were reviewed in phase four of the analysis through a cyclical process that involves back-and-forth movements between phases of data analysis until a consensus was reached on the final themes. Consequently, in phase five, "Defining and naming themes" was completed, through refining existing themes and subthemes that will be presented in the final analysis [34,35].

Results

Study Participants

Four FGDs were conducted among 32 women of reproductive age attending IRC clinics and Ministry of Health primary care facilities. In addition, 17 midwives aged between 27 and 57 years and with 1 month to 25 years of experience were interviewed. All midwives were previously trained to provide FP counseling and services.

Themes and Subthemes Categorization

The data generated through the focus groups and interviews were organized into 3 major themes, combining both the clients and providers' perspectives: (1) advantages of digital technology for FP, (2) concerns about the use of digital technology for FP, and (3) the characteristics of an ideal app for delivering an intervention on FP methods. Within the themes, 10 subthemes were generated. The themes and subthemes are summarized in Textbox 1.

Textbox 1. Themes (bold) and subthemes of midwives and women's perceptions about family planning in Jordan.

Advantages of digital technology for family planning

- Raising awareness on family planning
- Convenience and cost efficiency
- Empowering women

Concerns about the use of digital technology

- Quality and accuracy of the information
- eHealth literacy
- Need for advice from a health care professional
- Internet access and health literacy

Ideal app or website characteristics

- Ease of understanding and visualizations
- Privacy concerns
- Free access

Advantages of Digital Technology for Family Planning

Both clients and providers perceived digital technology as beneficial for FP for several reasons: it helps in raising awareness, is convenient and cost-effective, and may help in empowering women.

Raising Awareness on FP

Digital technology was perceived by the midwives and clients as an effective tool to increase the women and their husbands' awareness about FP. It was perceived to be effective in increasing their access to information about the different FP methods and their side effects, which will allow them to make better decisions on FP use. A midwife explained:

XSL•FO
RenderX

> *If a woman used a method and suffered from a side effect that the midwife did not talk about before, she can open an application and learn more about the side effect.*

A woman said, "It would be so helpful as I can take my time reading about each method." Moreover, some midwives believe that women tend to absorb information better when they search for it personally and they find this information more convincing. Another benefit of digital technology is that it could inform them about other aspects such as nutrition, physical, and mental health related to FP. A woman said: "I can gain information about nutrition, physical, and mental health related to FP."

Finally, digital technology was not only perceived as a tool to access information about FP for the patients, but midwives also perceived it as a source of new information for them. An interviewee explained "even if I'm a midwife, I download Daleel Hamly application (ie, my pregnancy guide) when I get pregnant, as I find some new information I've never studied before."

Convenience and Cost Efficiency

According to both providers and clients, digital tools were perceived as a convenient and cost-efficient method to promote FP as it can help women access FP-related information and answers their questions at any time based on their availability and free time. One client explained, "It would be great to obtain medical consultations through the websites while I am home, not worrying about leaving my children home alone."

Moreover, technology can help women to connect with health care providers without having to go to the health care center, especially for those who may not be able to seek FP services at the centers due to distance or the lack or inability to pay for transportation. A participant explained, "I live 30 minutes away from the health center and takes me time to find a taxi or bus and costs me at least 3 Jordanian dinars."

According to both providers and clients, technology can help women and their husbands decide about their preferred FP method before visiting the health care center, which may save both time and effort. A midwife said, "on the contrary, it would be useful if they were previously educated at home and they have already made up their minds before coming to the center to get the method." A woman explained, "My husband and I can reach mutual agreement and understanding of the preferred method we want to use before coming to the center to get the method as this will reduce consultation time."

Empowering Women

Most of the clients (Syrian refugees and Jordanian women) and providers perceived that digital technology can empower women because its information help women to convince their husbands to be involved in FP counseling sessions, especially if the husband cannot accompany women to counseling. Midwives explained that women can try to convince their husbands and in-laws of a certain method, especially if the husband cannot accompany women to counseling or if the midwives are embarrassed to explain the use of certain methods to the husband, by showing them a full explanation of it through a mobile app or website, as people tend to believe the information they find online. Clients and midwives stated that technology such as videos of teaching communication skills can teach women how to communicate with their husbands to seek FP services.

Additionally, midwives perceive that technological developments can increase the educated women's awareness about the available FP methods by enabling them to access a variety of topics easily and by giving them the chance to share and discuss their opinions and make their own decisions. Another reason was provided by Syrian and Jordanian women who mentioned that mobile technology can empower them through informing them of ways to take good care of their physical and mental health, how to space between pregnancies, and what general information each woman must know about the reproductive health.

Concerns About the Use of Digital Technology

Despite the high level of acceptance and enthusiasm toward the use of technology for the purpose of FP, some study participants raised a number of concerns related to the quality and accuracy of information, the ability to use technologies, the need for advice from a health care professional, limited internet access, and limited health literacy.

Quality and Accuracy of the Information

Women clients had concerns that FP apps may not have sufficient or updated information. A participant said, "I need to make sure that the information listed is correct and up to date." Another participant added, "There might be new types of FP methods or new types of pills and I don't know about." Furthermore, one of the biggest fears of some of the interviewed midwives is that the technological services may provide wrong information that is likely to scare women from seeking the FP methods (eg, rare side effects and impact of some methods on weight or infertility). In addition, some midwives were concerned that apps may not be developed in the right way, may not include sufficient information, or may not be monitored or supervised by experts. For example, they stated that the apps might include incorrect terminology or unclear sentences, especially those that are translated from one language to another, hence appearing untrustworthy.

eHealth Literacy

Both midwives and women raised the concern about sharing misinformation due to their inability to correctly understand and explain the concepts included in FP apps or websites. These problems can be linked to the eHealth literacy model [37], which entails 6 domains of literacy, including basic or functional literacy (ie, the ability to read a text); scientific, media, information (ie, the ability to understand scientific literature, how media work, and how information is shared and produced, respectively); health (ie, the ability to process and use health-related information on a specific topic to make decisions); and digital literacy (ie, the ability to use technology to seek information). eHealth literacy is closely linked to the level of education, which can be a barrier to interventions delivered through digital technology. Education plays a role especially among Syrian women, as, on average, their educational level is lower than that of Jordanian women. One Syrian participant

XSL•FO
RenderX

stated, "usually in Syria, girls get off married early before they complete their 8th grade," so they drop-out from school at an early age, not completing their education. This, of course, means that some of these women will not be able to read and understand some textual information. For example, both clients and providers mentioned the problems related to functional literacy: one woman mentioned that "I am worried that I will not understand what is written"; several providers expressed concerns about the readability of the text included in the apps, noting that they had observed that women can speak well but are not able to read. Some other women mentioned problems of scientific and health literacy that were linked to the difficulty of understanding some medical terms. Additionally, both midwives and women expressed concerns about their ability to use technology, as not all clients have familiarity with apps and websites.

Need for Advice From a Health Care Professional

Some midwives were hesitant about the use of technology in the FP-related services, as one thinks that women tend to take advice from each other, even if they are wrong. Some midwives think that technology cannot replace face-to-face consultations by the midwives, as women trust the health care providers and their professional experience. As for the clients, some believe that they might not know which type of FP method is best for their body without consulting their midwives, as one participant said, "How can I tell if this type is good for me without consulting the midwife."

In addition, 2 providers were concerned that digital technology apps would increase their workload by taking a longer time to counsel women. They believed women might have more questions about the many methods of FP after using apps, resulting in more time to counsel women on these additional methods.

Furthermore, it is important to know the medical history of each woman before giving her any advice regarding FP. A midwife working with Syrian women argued that technology may only provide a brief explanation of a certain medical condition, but not all of the smallest details and that it may not be able to warn the woman in case she needs an urgent medical intervention, including misunderstanding and information overlap, which may lead to making wrong decisions on FP. For example, a midwife working with Syrian women said, "Women may associate some of the side-effects presented through the technological services to the family planning method they use, even if the complications may be caused by other medical problems."

When women were asked about risks and consequences that might be associated with the use of digital technology, specifically about gender-based violence, they all stated that no harmful risks or consequences are associated with using digital technology for FP. However, midwives perceive that being well-educated about FP methods through technology may cause the woman a lot of problems with her husband, as she will use the information to justify her personal decisions, even if they are thoughtless or inconsiderate (eg, abortion).

Internet Access and Health Literacy

According to clients and midwives, most women have smartphones and internet service, and they use social media platforms all the time which makes it easier and faster for them to obtain family information; a participant explained, "we all have smart phones and access to the internet … we can ask questions online and get answers." Midwives perceive that women are curious in nature and they tend to ask questions and answer them, and they like to get medical consultations through the websites. For example, a midwife explained, "women tend to believe what technology tells them; when one asks a question through Facebook groups, a hundred reply with completely different answers. She then comes to the center telling me that I asked and got the answers myself ...." A midwife explained that, "Mobile phones are available for all people, and women living here like to know about the FP methods. In most cases, internet services are available; even if the woman's financial situation is not that good, she usually has a mobile phone and an internet connection." However, as one of the concerns, midwives and some women believe that the technological support may be limited due to technical errors or constant internet disconnection. A participant explained, "I worry that the internet cuts off during the time I need to have answers regarding family planning." Besides, a midwife at a rural health care center mentioned that technology may not be effective in the less developed regions. As people are still strict against using mobile devices at home, women may still not know how to use mobile technologies due to their low educational level or the impact of the surrounding society. It is also possible that some women may even be unable to afford owning a mobile phone. A midwife stated, "Some regions are far from the services, and they are still strict somehow."

Ideal App or Website Characteristics

The Jordanian and Syrian women and midwives described the ideal characteristics of a potential digital health app for FP. The needed and desired features mainly included visualizations and videos, being easy to understand, having a log-in feature (password protected) to maintain privacy, and free of charge accessibility.

Ease of Understanding and Visualizations

Most midwives thought that using digital health solutions for FP services will be highly and widely accepted by women under certain conditions. One major factor to take into consideration depends on the way of presenting and explaining information; a midwife explained that it will be more effective to use a simple language to be understandable by women of all educational and ideological levels. Most midwives reported that "video" is more effective than text as they believe some women are illiterate and they find videos catchier and more memorable. Some of the Jordanian and Syrian clients believed that information provided through videos is easier to understand and remember, especially for illiterate women, whereas others do not have issues with reading due to their high level of education. As we mentioned above, when talking about eHealth literacy, the level of education can be a barrier to interventions delivered through digital technology, especially among Syrian women, as, on average, their educational level is lower than that of Jordanian

XSL•FO
RenderX

women, as was stated by a Syrian participant "usually in Syria, girls get off married early before they complete their 8th grade," so they drop-out from school at an early age. Hence, it is important to make sure that low literacy is considered. In fact, one client mentioned, "The website of mobile application should be presented with lots of visualization and videos, which will be more effective to understand and apply." Another woman said, "I prefer watching a video because I am a visually oriented person." However, some clients thought it would be a good idea to provide the information as both text and video to fit everyone's taste. Some interviewees stated that videos are easier to use as they can pause or rewatch them when they are free, as noted by another participant, "I can watch the video any time I want and anywhere I am at." A midwife also said, "women can watch the video in their free time or even listen to it while doing other tasks at home." Some midwives added that women can show the videos to their husband or mother-in-law.

Privacy Concerns

When Jordanian and Syrian women were asked about the level of acceptance of digital technology for them and their husbands, all participants confirmed their acceptance. In fact, they stated that their husbands will not mind them accessing information online, if it is safe information, and no one can find out. This was noted by a participant: "my husband will make sure that no one except me and him know about using the mobile application."

Some clients preferred using password-protected platforms to protect their privacy and prevent their children (especially young daughters) opening the app and reading information that is not appropriate for their age. A participant said, "If there will be videos explaining – how to use FP methods – then they prefer to have a login account." Another participant explained, "I need to know that the application is secured by having a login account, so my daughter does not open it and read."

Free Access

Many midwives believed that it is fundamental that many clients accept and utilize such services so that they become cost-effective. Apps and websites need to be free of charge so that anyone can access them, as some midwives lamented the fact that some clients would not pay to get FP-related services due to financial limitations.

Discussion

Principal Findings

This study provides a significant contribution to knowledge as it is the first to explore the perceptions of women from refugee and host communities and midwives in Jordan toward the potential use of digital technology for enhancing FP. Overall, the response of the Syrian and Jordanian women and midwives was positive. According to most of the participating women in our study, it would be ideal if FP was discussed and provided at every level of digital technology, especially through mobile apps and websites (or web applications).

Advantages of Digital Technology

The study participants perceived that digital technology may serve as means to increase awareness and knowledge related to the different FP methods and their physical and mental side effects. In this regard, several previous studies have supported the use of digital technology in FP. For example, a study conducted by Greenleaf et al [38] showed that women who own a mobile phone were more likely to use modern FP methods compared with women who do not own one. Another study conducted in Bangladesh aimed to evaluate the effect of digital health package on FP knowledge and behavior through providing offline digital health training [39]. The findings showed improvement among women who received digital health package on FP regarding the choice of contraception, contraception method side effect, and management of side effects as compared with those without exposure at all. It was also found that prevalence of modern FP methods among women who received a digital health package on FP was high compared with those without exposure at all [39].

Another major advantage of digital technology is that it can aid the decision-making process and empower women. Almost all interviewed midwives and women (clients) agreed that digital and mobile technology could enable women and empower them to make informed decisions and can teach them how to make joint decisions with their husbands by following specific communication and bargaining skills and strategies. Several studies investigated the effect of digital and mobile technologies for FP on women empowerment and their influence on the decision-making process. A clustered randomized control trial conducted in Nigeria assessed the efficacy of the digital health tool 'Smart Client' on ideational and behavioral variables related to family [23]. The study researchers have found that the intervention group, which received regular mobile phone calls, showed improvements in the confidence level of women while discussing FP with a health care provider, and that women in this group tend to use modern FP methods more than those in the control group [23].

Concerns About Digital Technology for FP

When we asked the study participants whether the mobile technology would replace counseling services provided by midwives, the majority disagreed, but mentioned that it can be developed in conjunction with the available services provided at medical health centers. A study in Kenya indicated that mobile health (mHealth) alone was unable to improve contraception knowledge and use [40]. Nonetheless, previous studies concluded that appropriate use of the mobile technology needs biomedical screening and counseling, thus empowering selection of FP methods based on lifestyle and priorities of women [41].

Women are increasingly fascinated by using science-based mobile technology in contraception and family practices. However, most of the available mobile apps support natural FP methods [41]. A previous study aimed to evaluate smartphone apps that are designed to help users of FP methods to prevent unintended pregnancy [42]. Of the 218 apps identified in that study, 12 scored 50 out of 90 points on features and 15 out of 21 points on contraception and pregnancy prevention best practice. Moreover, 41% of the apps did not contain any

XSL•FO
RenderX

information about modern FP methods, and 21% contained information about 1 method only, with only half of the apps that briefed about modern contraception methods providing instructions for using FP methods [42]. Hence, digital health interventions should be aimed at modifications in behavior and shifts to new practices such as moving away from paper-based systems to digital methods [1]. Recently, the WHO has issued a guideline that provides 10 evidence-based recommendations on digital health interventions focusing on factors that influence the feasibility and acceptability of implementing digital health technologies, especially in low-resource settings, and taking into account gender, equity, and human rights issues [1]. Therefore, health care providers are encouraged to follow this useful WHO guideline when developing or implementing digital health interventions. At the same time, we believe that researchers should carefully design interventions taking into account the different levels and domains of literacy that pertain to the eHealth literacy model [37]. In our study, both clients and providers mentioned the need to provide content in a visual format, catering for participants with low literacy levels, and to make it available in an easy-to-use format. Curating content and making sure the information is accurate and correct are also fundamental elements of digital interventions for FP, which can avoid the diffusion of misinformation and disinformation (myths and wrong beliefs) about FP [43].

Ideal Apps or Websites

Based on the opinions shared by our study participants, mobile apps for FP should complement—not substitute—the information and services available, linking clients to providers. The information should be accurate, easy to understand, and verified by professionals. The apps or web services should be easy to use, based on videos or a combination of visual and textual information, accessible without the internet, and password protected. These services should be free of charge.

Strengths and Limitations

This qualitative study has several strengths. On one hand, the narratives included the perceptions of women and midwives to ensure triangulation of data and maximize the credibility and trustworthiness of the results. On the other hand, the researchers who conducted the data collection and analysis had vast experience in qualitative research methods. As for the limitations, as with most qualitative research designs, the study findings cannot be generalized to a wider population especially because we only included 2 governorates in Jordan. However, the study findings offer in-depth insights into personal and contextual factors affecting digital technology use for enhancing FP in Jordan.

Conclusions

As perceived by midwives and women, incorporating digital technology in FP services can be feasible, cost-effective, well accepted, and potentially beneficial in increasing woman's awareness regarding the FP methods and their side effect. It also may improve shared decision-making process among Jordanian couples. Therefore, digital technologies are recommended to address the challenges faced in health system and to improve both the coverage and quality of FP services and practices.

Acknowledgments

The authors thank the International Development Research Centre (IDRC) for their financial and technical support. This study is part of a 3-year project funded by the International Development Research Centre (IDRC), grant award: 109089-003.

Authors' Contributions

YK, MB, and MAN conceived and designed the study and YK coordinated it. MB and MAK provided intellectual feedback in the development of the study. NA-S and HY conducted the interviews, the focus groups discussion, transcribed the information, analyzed the transcripts, and developed the first set of themes. MAK and MB helped in the analyses and interpretation of the themes. YK, MAm, MAl drafted the manuscript; MAK and MB helped in the interpretation of the results and wrote the discussion. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Rodriguez MI, Kismodi E, Cottingham J, World Health Organization. Ensuring Human Rights in the Provision of Contraceptive Information and Services. Geneva, Switzerland: World Health Organization; 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK195054/
2. Machiyama K, Casterline JB, Mumah JN, Huda FA, Obare F, Odwe G, et al. Reasons for unmet need for family planning, with attention to the measurement of fertility preferences: protocol for a multi-site cohort study. Reprod Health 2017 Feb 09;14(1):23 [FREE Full text] [doi: 10.1186/s12978-016-0268-z] [Medline: 28183308]
3. Wulifan JK, Brenner S, Jahn A, De Allegri M. A scoping review on determinants of unmet need for family planning among women of reproductive age in low and middle income countries. BMC Womens Health 2016 Jan 15;16:2 [FREE Full text] [doi: 10.1186/s12905-015-0281-3] [Medline: 26772591]
4. World Health Organization. Unmet need for family planning Internet. URL: http://www.who.int/reproductivehealth/topics/family_planning/unmet_need_fp/en/ [accessed 2019-07-08]

XSL•FO
RenderX

5. Tanabe M, Myers A, Bhandari P, Cornier N, Doraiswamy S, Krause S. Family planning in refugee settings: findings and actions from a multi-country study. Confl Health 2017 May 31;11(1):9. [doi: 10.1186/s13031-017-0112-2]
6. UNFPA Jordan. Among Syrian refugees, dispelling myths about contraceptives. URL: https://jordan.unfpa.org/en/news/among-syrian-refugees-dispelling-myths-about-contraceptives [accessed 2019-06-19]
7. Department of Statistics (DOS), ICF. Jordan Population and Family Health Survey 2017-18 Internet. Amman, Jordan and Rockville, Maryland, USA: DOS and ICF; 2019. URL: https://dhsprogram.com/pubs/pdf/FR346/FR346.pdf [accessed 2021-07-15]
8. Sieverding M, Berri N, Abdulrahim S. Marriage and fertility patterns among Jordanians and Syrian refugees in Jordan (Report No.: 1187). 2018 May. URL: https://ideas.repec.org/p/erg/wpaper/1187.html [accessed 2021-07-15]
9. Situation Syria Regional Refugee Response. URL: https://data2.unhcr.org/en/situations/syria/location/71 [accessed 2021-07-15]
10. Chandra-Mouli V, McCarraher DR, Phillips SJ, Williamson NE, Hainsworth G. Contraception for adolescents in low and middle income countries: needs, barriers, and access. Reprod Health 2014 Jan 02;11(1):1 [FREE Full text] [doi: 10.1186/1742-4755-11-1] [Medline: 24383405]
11. Jordan Communication, Advocacy and Policy Activity (JCAP). Exploring Gender Norms and Family Planning in Jordan: A Qualitative Study. USAID|Jordan. 2016 Jan. URL: https://jordankmportal.com/resources/exploring-gender-norms-and-family-planning-in-jordan [accessed 2021-07-17]
12. Newman K, Feldman-Jacobs C. Family Planning and Human Rights—What's the Connection and Why Is It Important? Population Reference Bureau. 2015 Jul. URL: https://www.prb.org/resources/family-planning-and-human-rights-whats-the-connection-and-why-is-it-important/ [accessed 2021-07-17]
13. West L, Isotta-Day H, Ba-Break M, Morgan R. Factors in use of family planning services by Syrian women in a refugee camp in Jordan. J Fam Plann Reprod Health Care 2016 Apr 11;73(2):96-102 [FREE Full text] [doi: 10.1136/jfprhc-2014-101026] [Medline: 26962045]
14. Bitzer J, Abalos V, Apter D, Martin R, Black A, Global CARE (Contraception: Access, Resources, Education) Group. Targeting factors for change: contraceptive counselling and care of female adolescents. Eur J Contracept Reprod Health Care 2016 Dec;21(6):417-430. [doi: 10.1080/13625187.2016.1237629] [Medline: 27701924]
15. Department of Statistics. Population Projections for the Kingdom's Residents during the Period 2015-2050. Department of Statistics. 2016 Dec. URL: http://www.dos.gov.jo/dos_home_e/main/Demograghy/2017/POP_PROJECTIONS(2015-2050).pdf [accessed 2021-07-17]
16. Hameed W, Azmat SK, Ali M, Sheikh MI, Abbas G, Temmerman M, et al. Women's empowerment and contraceptive use: the role of independent versus couples' decision-making, from a lower middle income country perspective. PLoS One 2014;9(8):e104633 [FREE Full text] [doi: 10.1371/journal.pone.0104633] [Medline: 25119727]
17. Halpern V, Grimes DA, Lopez L, Gallo MF. Strategies to improve adherence and acceptability of hormonal methods for contraception. Cochrane Database Syst Rev 2006 Jan 25(1):CD004317. [doi: 10.1002/14651858.CD004317.pub2] [Medline: 16437483]
18. Lopez L, Grey T, Tolley E, Chen M. Brief educational strategies for improving contraception use in young people. Cochrane Database Syst Rev 2016 Mar 30;3:CD012025 [FREE Full text] [doi: 10.1002/14651858.CD012025.pub2] [Medline: 27027480]
19. Car J, Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2012 Jul 11(7):CD007458. [doi: 10.1002/14651858.CD007458.pub2] [Medline: 22786507]
20. Schwebel FJ, Larimer ME. Using text message reminders in health care services: A narrative literature review. Internet Interv 2018 Sep;13:82-104 [FREE Full text] [doi: 10.1016/j.invent.2018.06.002] [Medline: 30206523]
21. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
22. McCarthy OL, Wazwaz O, Osorio Calderon V, Jado I, Saibov S, Stavridis A, et al. Development of an intervention delivered by mobile phone aimed at decreasing unintended pregnancy among young people in three lower middle income countries. BMC Public Health 2018 May 02;18(1):576 [FREE Full text] [doi: 10.1186/s12889-018-5477-7] [Medline: 29716571]
23. Babalola S, Loehr C, Oyenubi O, Akiode A, Mobley A. Efficacy of a Digital Health Tool on Contraceptive Ideation and Use in Nigeria: Results of a Cluster-Randomized Control Trial. Glob Health Sci Pract 2019 Jun;7(2):273-288 [FREE Full text] [doi: 10.9745/GHSP-D-19-00066] [Medline: 31249023]
24. Agarwal S, Lasway C, L'Engle K, Homan R, Layer E, Ollis S, et al. Family Planning Counseling in Your Pocket: A Mobile Job Aid for Community Health Workers in Tanzania. Glob Health Sci Pract 2016 Jun 20;4(2):300-310 [FREE Full text] [doi: 10.9745/GHSP-D-15-00393] [Medline: 27353622]
25. Braun R, Lasway C, Agarwal S, L'Engle K, Layer E, Silas L, et al. An evaluation of a family planning mobile job aid for community health workers in Tanzania. Contraception 2016 Jul;94(1):27-33. [doi: 10.1016/j.contraception.2016.03.016] [Medline: 27039033]

XSL•FO
RenderX

26. Diedhiou A, Gilroy KE, Cox CM, Duncan L, Koumtingue D, Pacqué-Margolis S, et al. Successful mLearning Pilot in Senegal: Delivering Family Planning Refresher Training Using Interactive Voice Response and SMS. Glob Health Sci Pract 2015 Jun 02;3(2):305-321 [FREE Full text] [doi: 10.9745/GHSP-D-14-00220] [Medline: 26085026]
27. Alencar A, Kondova K, Ribbens W. The smartphone as a lifeline: an exploration of refugees' use of mobile communication technologies during their flight. Media, Culture & Society 2018 Nov 30;41(6):828-844. [doi: 10.1177/0163443718813486]
28. Narli N. Life, Connectivity and Integration of Syrian Refugees in Turkey: Surviving through a Smartphone. questionsdecommunication 2018 Sep 01(33):269-286. [doi: 10.4000/questionsdecommunication.12523]
29. About the Global Digital Health Index. Global Digital Health Index. 2020. URL: https://www.digitalhealthindex.org/ [accessed 2021-07-17]
30. Smith C, Ly S, Uk V, Warnock R, Free C. Women's views and experiences of a mobile phone-based intervention to support post-abortion contraception in Cambodia. Reprod Health 2017 Jun 12;14(1):72 [FREE Full text] [doi: 10.1186/s12978-017-0329-y] [Medline: 28606181]
31. Biswas KK, Hossain A, Chowdhury R, Andersen K, Sultana S, Shahidullah SM, et al. Using mHealth to Support Postabortion Contraceptive Use: Results From a Feasibility Study in Urban Bangladesh. JMIR Form Res 2017 Oct 27;1(1):e4 [FREE Full text] [doi: 10.2196/formative.5151] [Medline: 30684398]
32. Mboane R, Bhatta MP. Influence of a husband's healthcare decision making role on a woman's intention to use contraceptives among Mozambican women. Reprod Health 2015 Apr 23;12:36 [FREE Full text] [doi: 10.1186/s12978-015-0010-2] [Medline: 25902830]
33. Kriel Y, Milford C, Cordero J, Suleman F, Beksinska M, Steyn P, et al. Male partner influence on family planning and contraceptive use: perspectives from community members and healthcare providers in KwaZulu-Natal, South Africa. Reprod Health 2019 Jun 25;16(1):89 [FREE Full text] [doi: 10.1186/s12978-019-0749-y] [Medline: 31238960]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Clarke V, Braun V. Thematic analysis. The Journal of Positive Psychology 2016 Dec 09;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
36. van Nes F, Abma T, Jonsson H, Deeg D. Language differences in qualitative research: is meaning lost in translation? Eur J Ageing 2010 Dec;7(4):313-316 [FREE Full text] [doi: 10.1007/s10433-010-0168-y] [Medline: 21212820]
37. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
38. Greenleaf AR, Ahmed S, Moreau C, Guiella G, Choi Y. Cell phone ownership and modern contraceptive use in Burkina Faso: implications for research and interventions using mobile technology. Contraception 2019 Mar;99(3):170-174 [FREE Full text] [doi: 10.1016/j.contraception.2018.11.006] [Medline: 30468721]
39. J. Limaye R, Kapadia-Kundu N, Arnold R, Gergen J, M. Sullivan T. Utilizing digital health applications as a means to diffuse knowledge to improve family planning outcomes in Bangladesh. Clin Obstet Gynecol Reprod Med 2017;3(2):1000176. [doi: 10.15761/cogrm.1000176]
40. Lee S, Begley C, Morgan R, Chan W, Kim SY. Addition of mHealth (mobile health) for family planning support in Kenya: disparities in access to mobile phones and associations with contraceptive knowledge and use. Int Health 2019 Nov 13;11(6):463-471. [doi: 10.1093/inthealth/ihy092] [Medline: 30576546]
41. Peterson SF, Fok WK. Mobile technology for family planning. Curr Opin Obstet Gynecol 2019 Dec;31(6):459-463. [doi: 10.1097/GCO.0000000000000578] [Medline: 31573996]
42. Mangone ER, Lebrun V, Muessig KE. Mobile Phone Apps for the Prevention of Unintended Pregnancy: A Systematic Review and Content Analysis. JMIR Mhealth Uhealth 2016;4(1):e6 [FREE Full text] [doi: 10.2196/mhealth.4846] [Medline: 26787311]
43. Sharma V, De Beni D, Sachs Robertson A, Maurizio F. Why the Promotion of Family Planning Makes More Sense Now Than Ever Before? Journal of Health Management 2020 Aug 05;22(2):206-214. [doi: 10.1177/0972063420935545]

Abbreviations

FGDs: focus group discussions
FP: family planning
GDHI: Global Digital Health Index
IDRC: International Development Research Center
IRC: International Rescue Committee
LMICs: low- and middle-income countries
mHealth: mobile Health
RQDA: Qualitative Data Analysis in R
UNHCR: United Nations High Commissioner for Refugees
WHO: World Health Organization

XSL•FO
RenderX

Edited by R Kukafka; submitted 22.11.20; peer-reviewed by M Khatatbeh, H Mehdizadeh; comments to author 29.12.20; revised version received 08.01.21; accepted 24.05.21; published 28.07.21.

Please cite as:
Yousef H, Al-Sheyab N, Al Nsour M, Khader Y, Al Kattan M, Bardus M, Alyahya M, Taha H, Amiri M
Perceptions Toward the Use of Digital Technology for Enhancing Family Planning Services: Focus Group Discussion With Beneficiaries and Key Informative Interview With Midwives
J Med Internet Res 2021;23(7):e25947
URL: https://www.jmir.org/2021/7/e25947
doi:10.2196/25947
PMID:34319250

©Hind Yousef, Nihaya Al-Sheyab, Mohannad Al Nsour, Yousef Khader, Malika Al Kattan, Marco Bardus, Mohammad Alyahya, Hana Taha, Mirwais Amiri. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measurement of Cancer-Related Fatigue Based on Heart Rate Variability: Observational Study

Chi-Huang Shih[1], PhD; Pai-Chien Chou[2,3], MD, PhD; Ting-Ling Chou[4], MSc; Tsai-Wei Huang[4,5,6], PhD

[1]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan

[2]Division of Thoracic Medicine, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[3]Division of Thoracic Medicine, Department of Internal Medicine, School of Medicine, Taipei Medical University, Taipei, Taiwan

[4]School of Nursing, College of Nursing, Taipei Medical University, Taipei City, Taiwan

[5]Cochrane Taiwan, Taipei Medical University, Taipei, Taiwan

[6]Center for Nursing and Healthcare Research in Clinical Practice Application, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Tsai-Wei Huang, PhD
School of Nursing
College of Nursing
Taipei Medical University
250, Wuxing Street,
Taipei City, 11031
Taiwan
Phone: 886 2 27361661 ext 6350
Fax: 886 2 3772842
Email: tsaiwei@tmu.edu.tw

Abstract

Background: Cancer-related fatigue is a serious side effect of cancer, and its treatment can disrupt the quality of life of patients. Clinically, the standard method for assessing cancer-related fatigue relies on subjective experience retrieved from patient self-reports, such as the Brief Fatigue Inventory (BFI). However, most patients do not self-report their fatigue levels.

Objective: In this study, we aim to develop an objective cancer-related fatigue assessment method to track and monitor fatigue in patients with cancer.

Methods: In total, 12 patients with lung cancer who were undergoing chemotherapy or targeted therapy were enrolled. We developed frequency-domain parameters of heart rate variability (HRV) and BFI based on a wearable-based HRV measurement system. All patients completed the BFI-Taiwan version questionnaire and wore the device for 7 consecutive days to record HRV parameters such as low frequency (LF), high frequency (HF), and LF-HF ratio (LF-HF). Statistical analysis was used to map the correlation between subjective fatigue and objective data.

Results: A moderate positive correlation was observed between the average LF-HF ratio and BFI in the sleep phase (ρ=0.86). The mapped BFI score derived by the BFI mapping method could approximate the BFI from the patient self-report. The mean absolute error rate between the subjective and objective BFI scores was 3%.

Conclusions: LF-HF is highly correlated with the cancer-related fatigue experienced by patients with lung cancer undergoing chemotherapy or targeted therapy. Beyond revealing fatigue levels objectively, continuous HRV recordings through the photoplethysmography watch device and the defined parameters (LF-HF) can define the active phase and sleep phase in patients with lung cancer who undergo chemotherapy or targeted chemotherapy, allowing a deduction of their sleep patterns.

(J Med Internet Res 2021;23(7):e25791) doi:10.2196/25791

KEYWORDS

cancer-related fatigue; heart rate variability; LF-HF ratio; photoplethysmography; wearables; chemotherapy

XSL•FO
RenderX

Introduction

Background

Fatigue is a reminder of the requirement of energy reimbursement in normal individuals. Cancer-related fatigue, defined as a multidimensional phenomenon that develops over time, is characterized by diminished energy and mental capacity and disturbed psychological conditions among patients with cancer [1]. The prevalence of cancer-related fatigue is estimated to vary from 60% to 90%, depending on the diagnostic criteria used [2]. Cancer-related fatigue often coexists with symptoms of depression, pain, anorexia, insomnia, anxiety, nausea, and dyspnea, all of which can contribute to the expression of fatigue [3]. Cancer-related fatigue imposes limitations on the normal daily activities of several patients with cancer and profoundly affects all aspects of their quality of life, including compliance with standard treatments such as chemotherapy or radiotherapy [4]. Thus, cancer-related fatigue management is an essential part of the treatment plan for patients with cancer.

To evaluate the severity of cancer-related fatigue, patients in most studies were subjectively evaluated through questions such as “How would you rate your fatigue on a scale of 0-10 over the past 7 days?” (0=no fatigue; 10=worst fatigue you can imagine). Additional measurements could be conducted using other standardized assessment tools, such as the Brief Fatigue Inventory (BFI). According to the guideline’s recommendation, regular screening of fatigue in clinical practice has been suggested [5]. However, cancer-related fatigue often remains underreported and untreated for a variety of reasons [2]. First, most of the methods used to measure cancer-related fatigue are subjective. Second, patients may gradually become accustomed to the impaired physical condition and consider discomfort as normal. Third, although several drugs have been used in clinical practice to relieve cancer-related fatigue, no standard treatments thus far have been recommended in clinical guidelines or proved effective.

Heart rate variability (HRV) is measured using a method similar to electrocardiogram (ECG) because HRV indicates the variability between two heartbeats. In practice, HRV signals are obtained from ECGs of at least 240 seconds [6] and heart rate (HR) recovery from maximal exercise testing [7]. HRV is continuously modulated through complex interactions by the autonomic nervous system, involving the sympathetic nervous system (SNS), parasympathetic nervous system (PNS), and vagus nerve [8]. SNS activities increase the HR, whereas those of the PNS reduce it [9,10]. HRV has two parameters: the time domain and frequency domain. Frequency-domain HRV is widely used to detect fatigue and its corresponding effects, such as drowsiness and sleep. Current frequency-domain measurements assign the bands of frequency into a high frequency (HF) range between 0.14 and 0.4 Hz and a low frequency (LF) range between 0.05 and 0.15 Hz. The LF-HF ratio (LF-HF) represents the relative activity between the SNS and PNS under controlled conditions [11]. A decrease in the HF band and an increase in the LF band characterize low parasympathetic activity. Previous studies have suggested that stress affects the SNS, PNS, and vagus nerve activities; thus, HRV may be used to indicate the psychological health status of patients [12,13]. HRV is widely used to measure tiredness in different scenarios and diverse disease groups [14] and has been adopted to identify the awake state and fatigue state [15]. Although not clearly understood, current evidence indicates that a higher norepinephrine level in the brain might be the core of the symptoms [16]. The HF increases during the fatigue state, whereas the LF increases during the awake state. In the non–rapid eye movement (non-REM) state, the LF-HF ratio gradually decreases as sleep deepens [15,17]. The LH-HF ratio reflects sleep activity. LF-HF differentiates the non-REM sleep from the REM state. Sleep quality at night and rest frequency during daytime can potentially correspond to a fatigue status. A higher fatigue level may cause a higher resting frequency during the daytime, affecting sleep quality at night.

Sleep activity consists of REM and non-REM states. The non-REM state consists of stages 1, 2, 3, and 4, in which a higher stage implies deeper sleep. Dreams typically occur during the REM state, and the brain becomes more active than in the non-REM state. A sleeping cycle initiates with the non-REM state, starting from non-REM stages 1-4 and back to non-REM stage 1, followed by the REM state, before the non-REM state appears again. Each cycle lasts for approximately 90-120 minutes in an average adult [18]. However, LF-HF bursts may occur in the REM state, resulting in a much higher LF-HF ratio than in the non-REM state [19-21].

Objectives

In this study, we aim to develop objective cancer-related fatigue criteria based on HRV. We obtained the LF-HF ratio from HRV data collected from a wearable device with photoplethysmography (PPG) sensors to achieve this goal. The wearable device was built to collect HRV signals with a predefined frequency and to calculate the LF-HF ratio. The measurement was triggered per hour within 24 hours for 7 consecutive days.

Methods

HRV Measurement

HRV signals were contained in the data collection process during the ECG measurement (calculated). HRV is usually recorded every 5 minutes for up to 24 hours. Thus, HRV can be calculated from a similar device if the raw data can be extracted; this device has been used for overnight monitoring of ECG and HRV in cardiovascular assessments and studies [22,23]. However, lead-based ECG and HRV devices are designed for stationary measurements but not for mobile measurements, which are required in many physiological measurements. The continuous interpulse interval can be accumulated and converted into a wave, and the R-R intervals are measured and averaged to obtain information regarding the LF and HF.

PPG is a noninvasive technology that uses a light source and a photodetector at the surface of the patient’s skin to measure the volumetric variations of blood circulation and monitor personal health conditions [24]. Among the mobile designs of PPG, the wrist-worn design has become a commercialized device for

XSL•FO
RenderX

revealing general health indicators calculated from the volumetric variations of blood circulation or for research that requires these raw data from the patients for further analysis [25]. PPG signals offer an excellent substitute for ECG recordings [24]. A previous study compared the pulse rate variability obtained from PPG with HRV information collected from a classic stationary device worn by normal individuals and found significant correlations of >82% for both time and frequency features [25]. Thus, pulse rate variability indices can be used as surrogates for HRV indices.

In this study, we used a PPG smart band device developed by a local company (ViPCare, Gadgletech) to collect PPG information from the patients. The cloud system of this device converts PPG information into common HRV parameters, which can be commonly applied for diagnostic and research purposes. Each HRV measurement required 2 minutes to obtain the HR and involved two frequency-domain parameters, namely LF and HF. The converted HF and LF information was used for further statistical analysis to estimate the sympathetic and parasympathetic activities of the patients to assess their stress and fatigue [11].

Participants

The study protocol was approved by the Joint Institutional Review Board of Taipei Medical University (approval number N201910036). By using a convenient sampling approach, patients with lung cancer who visited the thoracic medicine clinic or those admitted to the thoracic medicine ward for chemotherapy or targeted chemotherapy in the teaching hospital of the Taipei Medical University were approached for potential enrollment. The inclusion criteria were ages 20 years and the ability to wear the PPG watch device. Patients with weak consciousness or those unable to respond to the questionnaire were excluded. Finally, 12 patients were included in the study.

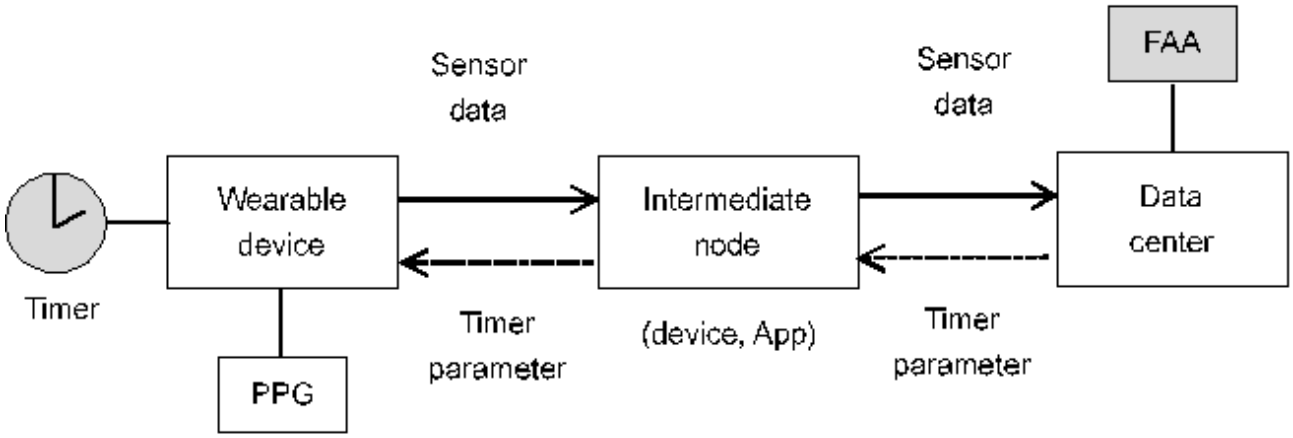
Sampling Procedures

The aim and purpose of the study were explained to the patients, and written informed consent was obtained from them before the investigation. The patients were asked to fill the BFI, Taiwan version (BFI-T), at baseline and were instructed to wear the PPG watch device immediately after finishing their questionnaire. The participants wore the PPG watch device continuously for 7 days, except while bathing. They were reminded by telephone to ensure that the device was worn at all times during the research period. After 7 days, the devices were collected from the participants during their next visit or admission or returned by the courier.

Measurement System Overview

We aimed to estimate the cancer-related fatigue index based on the HRV signals of the patients. To effectively observe the changes in the cancer-related fatigue during the daily lives of the patients, we used a wearable measuring device consisting of PPG sensors to monitor the HRV signals of the patients continuously for several days (Figure 1). The device consisted of a timer that triggered PPG sensors to collect timed HRV signals transmitted to an intermediate node via Bluetooth and forwarded them to the data center through Wi-Fi or ethernet for the fatigue analysis agent for further calculations. The data center communicated with the wearables to configure the timer parameters: measurement frequency (eg, 1 hour) and duration (eg, 2 min). Cancer-related fatigue analysis procedures, such as the timed sensor data, included the HRV parameters with their timestamps. According to the timestamp, HRV parameters can be classified into different phases. In this study, we considered the active phase and sleep phase as daytime and nighttime, respectively, for the patients. The resultant parameters included the mean and variance of LF, HF, and LF-HF. The fatigue analysis agent merged the continuous HRV parameters and conducted an analysis using the BFI.

Figure 1. A wearable-based heart rate variability measurement system. FAA: fatigue analysis agent; PPG: photoplethysmography.



XSL•FO
RenderX

Statistical Analysis

IBM SPSS Statistics (Version 23.0, IBM Corporation) was used for statistical analyses. Descriptive statistics, including percentage, mean, and SD, were used to present the general features of the data. Multivariate regression analysis was used to estimate the contributions of BFI to LF, HF, and LF-HF. We examined the data clusters and concerned the BFI definition (mild fatigue: <3; moderate fatigue: 4-7; severe fatigue: >7) to determine patient groups in the regressions. We used two engineering indexes, mean absolute error (MAE) and root mean square error (RMSE), to show the difference of error in paired observed values of HRV outcomes. The MAE shows the difference of error between the subjective and objective measurements whereas the RMSE represents the SD of the difference between predicted and observed values.

Results

Patient Characteristics

The descriptive statistics of the 12 patients are summarized in Table 1. The mean age of the patients was 65.3 years (SD 6.2), and the male:female ratio was 5:7. Most patients had a primary diagnosis of lung adenocarcinoma (10/12, 83%), and the rest had a primary cancer diagnosis of lung squamous cell carcinoma. Most patients were at an advanced stage of cancer (IV; 9/12, 75%). Targeted therapy with epidermal growth factor receptor–tyrosine kinase inhibitors was primarily used. Half of the patients were also treated with cardiovascular drugs. Most patients were also administered hypnotics to help them sleep, and most experienced some discomfort but were almost fully ambulant during the research period (Eastern Cooperative Oncology Group Performance Status=1).

Table 1. Descriptive statistics (N=12).

Variable	Values
Age (years), mean (SD)	65.3 (6.2)
Sex, n (%)	
Female	5 (42)
Male	7 (58)
Primary diagnosis, n (%)	
Lung adenocarcinoma	10 (83)
Lung squamous cell carcinoma	2 (17)
Stage, n (%)	
Ⅰ	1 (8)
Ⅲ	2 (17)
Ⅳ	9 (75)
Treatment, n (%)	
Target (EGFR-TKI[a])	9 (75)
Chemotherapy	3 (25)
Cardiovascular drugs, n (%)	
No	6 (50)
Yes	6 (50)
Hypnotics, n (%)	
No	10 (83)
Yes	2 (17)
ECOG[b], n (%)	
1	9 (75)
2	3 (25)

[a]EGFR-TKI: epidermal growth factor receptor–tyrosine kinase inhibitor.

[b]ECOG: Eastern Cooperative Oncology Group Performance Status.

Table 2 shows the results of the subjective and objective measurements of this study obtained using the PPG watch device and self-reports of BFI-T. The results of objective outcomes included HF and LF-HF data for 7 continuous days (24 hours) during the sleep period and active period of the patients. For patients who self-reported no, mild, or moderate levels of fatigue, the frequency of LF-HF >1 during the sleep period ranged from 0% to 25%, 36.7% to 44.8%, and 80.2% to 90.7%, respectively; the mean (SD) of LF-HF ranged from 0.73 (SD 0.07) to 1.14 (SD 0.51), 1.04 (SD 0.44) to 2.12 (SD 1.65), and

XSL•FO
RenderX

2.49 (SD 1.78) to 3.19 (SD 2.29), respectively. The values of these indicators increased with the BFI-T score.

Table 2. Subjective and objective statistics.

Participant ID	Objective			Subjective	
	Sleeping phase		Active phase; LF-HF[a]<1 (n=14), n (%)	BFI-T[b] score	Note
	LF-HF>1 (n=10), n (%)	LF-HF, mean (SD)			
P03	0 (0)	0.73 (0.07)	4.5 (32)	0	No
P13	0.8 (8)	0.91 (0.44)	7 (50)	0	No
P06	1.7 (17)	0.95 (0.20)	2.6 (19)	0	No
P02	2.5 (25)	1.14 (0.51)	8.6 (61)	0	No
P10	3.7 (37)	1.04 (0.44)	1.5 (11)	1.22	Mild
P07	3.3 (33)	1.04 (0.25)	5 (36)	1.33	Mild
P11	3.6 (36)	1.11 (0.33)	12.3 (88)	1.78	Mild
P01	4.5 (45)	1.11 (0.27)	0.2 (1)	1.56	Mild
P12	6 (60)	1.51 (0.73)	1.8 (13)	1.22	Mild
P04	4.5 (45)	2.12 (1.65)	1.5 (11)	1.11	Mild
P08	8 (80)	2.49 (1.78)	1.2 (9)	5	Moderate
P05	9.1 (91)	3.19 (2.29)	1.9 (14)	5.67	Moderate

[a]LF-HF: low frequency–high frequency ratio.

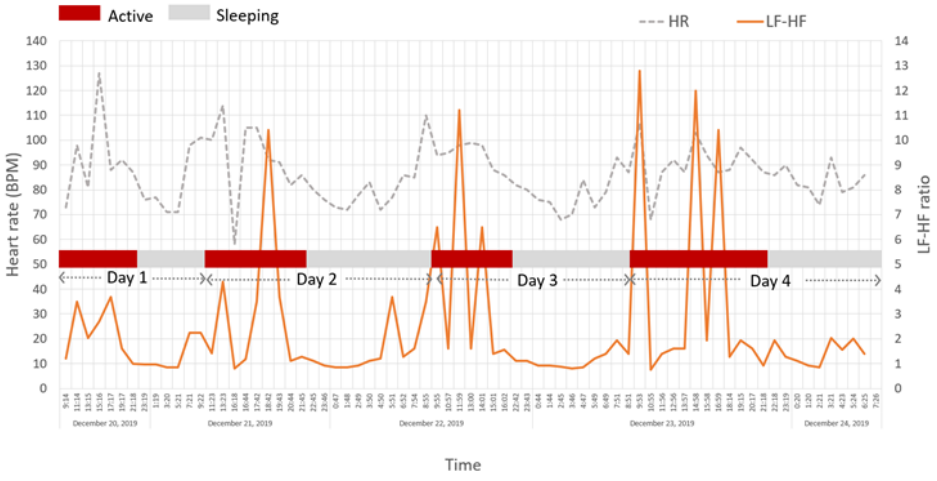
[b]BFI-T: Brief Fatigue Inventory-Taiwan version.

HRV-Based Fatigue Analysis

We used HR and LF-HF for the fatigue analysis in this study. LF-HF ratios are typically around 1 in the non-REM state and are even lower for a higher non-REM stage, whereas the LF-HF ratio increases in the REM state [19,20].

Figure 2 shows the real-time HR and LF-HF ratio traces of patient P01. Typically, a high activity level leads to an increased HR and LF-HF ratio and attains a higher average HR and LF-HF ratio in the active phases. In the sleep phase, the HR decreases gradually and the LF-HF ratio decreases from >1 to <1, and accordingly, improved sleep quality. Both the HR and LF-HF ratio increased at the end of the sleep phase. Sleep events also occurred during the active phase of the patient. A low LF-HF ratio (ie, <1) with a low HR occurs on days 2 and 4. Moreover, there is a burst in the LF-HF ratio curve, which indicates an REM state in the sleep phase on day 2. HRV is monitored in a timed-measurement random sampling manner, in which HRV events are randomly sampled in a predefined interval. To precisely capture the target events, we can adopt a fine-grained measurement interval in the timer configuration. Owing to the need for a large data storage space and electric power consumption, we set the measurement time to 1 hour.

Figure 2. Real-time heart rate variability tracing over a duration of 4 days for patient P01. HR: heart rate; HF: high frequency; LF: low frequency; BPM: beats per minute.



For patient *P*, we defined the LF-HF disorder ratio in the sleep phase as follows:

XSL•FO
RenderX

This ratio tracks the relationship between sleep quality and cancer-related fatigue. A higher value of  implies a shorter deep sleep duration. The LF-HF disorder ratio in the active phase for patient *P* can be calculated as follows:

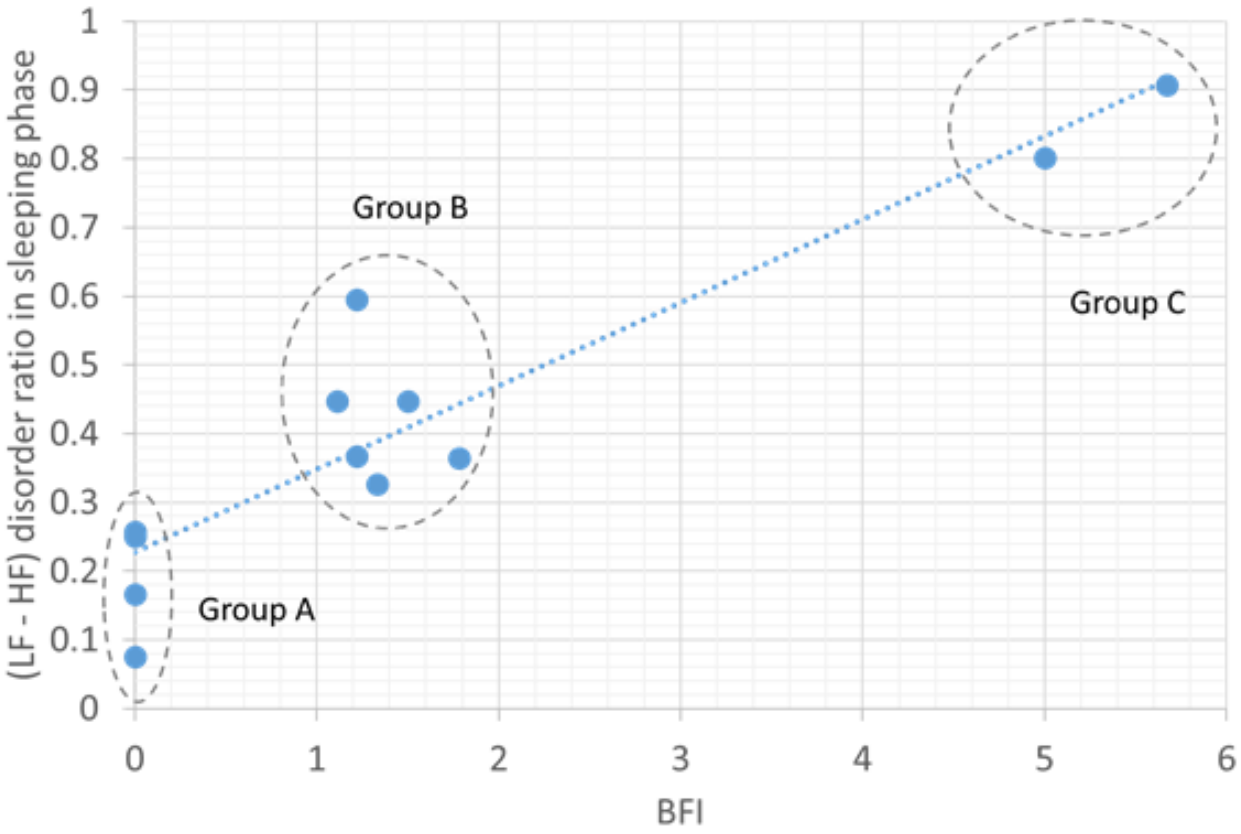

This ratio primarily tracks the fatigue level during daytime rest. A higher value of " represents more frequent sleep events during daytime.

Figures 3-5 show the correlations between objective HRV indicators and subjective BFI. Figure 3 demonstrates three facts: (1) the LF-HF disorder ratio increases with BFI in the sleep phase, (2) a strong positive correlation between the two axes (ρ=0.93), and (3) three distinct BFI groups, namely groups A (BFI=0), B (BFI=1-2), and C (BFI>3), exist between LF-HF disorder ratios of 0-1.

A moderate positive correlation was found between the average LF-HF and BFI in the sleep phase (ρ=0.86; Figure 4). The BFI groups for the LF-HF disorder ratio were distributed in a scattered manner and demonstrated a weak negative correlation in the active phase (ρ=−0.47; Figure 5). Thus, sleep quality is highly related to cancer-related fatigue, as measured by the BFI (Figures 3 and 4). Patients with high BFI (ie, high fatigue level) had poor sleep quality at night and insufficient daytime rest (Figures 3-5). The sleep quality of patients with lower BFI (ie, lower fatigue level) is generally better at night, and the resting frequency during daytime varies.

Figure 3. Relationship between the Brief Fatigue Inventory and low frequency or high frequency disorder ratio in the sleeping phase. The blue dotted line shows positive correlation. BFI: Brief Fatigue Inventory; HF: high frequency; LF: low frequency.



XSL•FO
RenderX

Figure 4. Relationship between the Brief Fatigue Inventory and average low frequency or high frequency ratio in the sleeping phase. The blue dotted line shows positive correlation. BFI: Brief Fatigue Inventory; HF: high frequency; LF: low frequency.

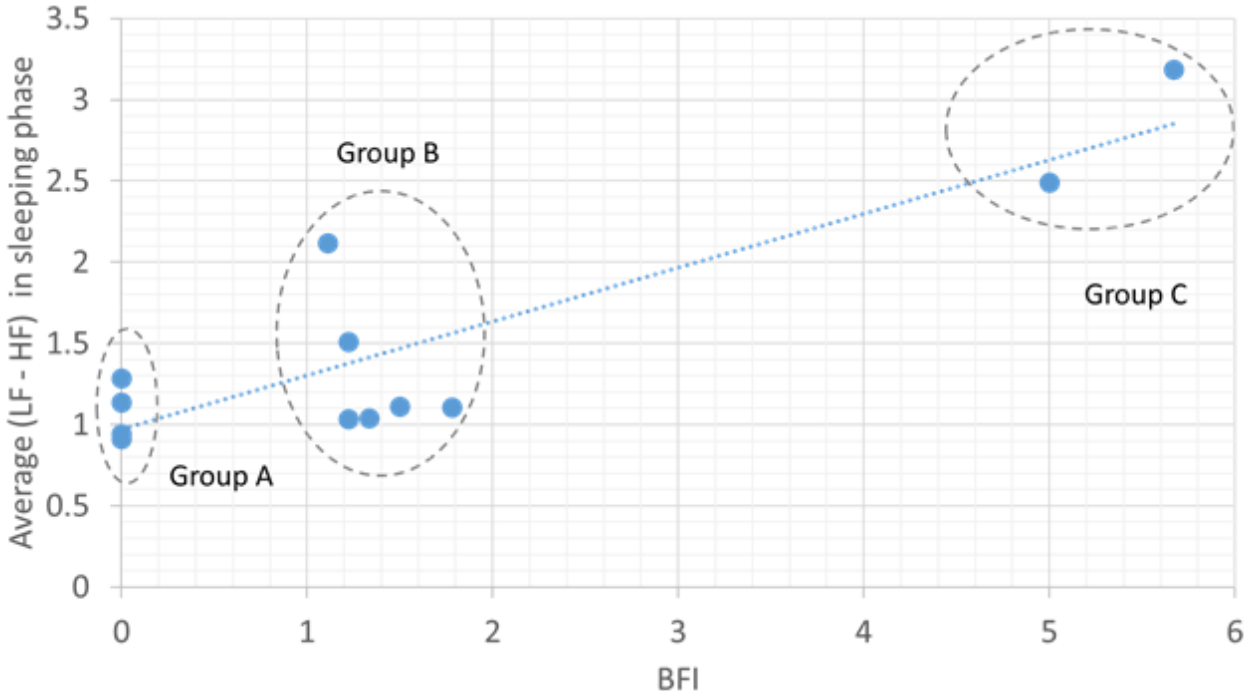
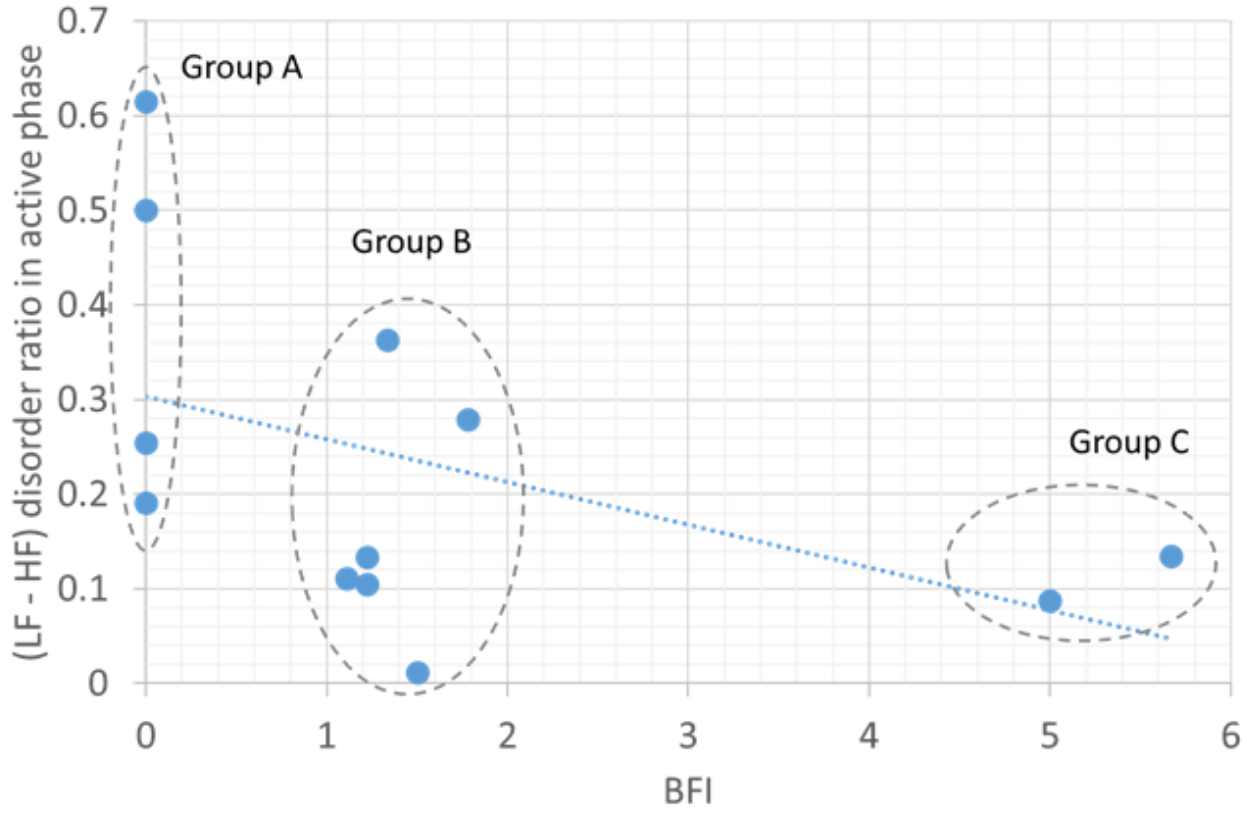


Figure 5. Relationship between the Brief Fatigue Inventory and low frequency or high frequency disorder ratio in the active phase. The blue dotted line shows negative correlation. BFI: Brief Fatigue Inventory; HF: high frequency; LF: low frequency.



BFI Mapping

In HRV-based BFI mapping, two HRV measurement factors need to be considered: (1) the correlation between HRV parameters and BFI and (2) the grouping effect. We defined the BFI mapping equation as follows:



XSL•FO
RenderX

where  represents the average LF-HF ratio in the sleep phase, ε stands for the shifting factor, and $MBFI_P$ stands for the mapped BFI for patient *P*. In equation (3), three weighting factors, namely α, β, and γ, are used to amplify or reduce the corresponding HRV parameters, depending on their correlation with the BFI, and their respective value ranges contributed to the BFI. While fitting equation (3) into a multiple linear regression model, the parameter ε corresponds to the intercept of the model, and based on the available data set of this study, the parameter vector {α, β, γ, ε} is {6.4, 0.45, 1.23, −2.07} with R^2=0.87.

When calculating the mapped BFI (MBFI), the optimal vectors {α, β, γ, ε} for each group can be retrieved from the statistical data, and the corresponding vector varies as the data set changes. To develop a general BFI mapping method without a grouping effect, we considered a common weighting vector {α=1, β=1, γ=1} to adapt the differences observed for groups and individuals. For patient *P*, the difference between the MBFI values is shown in equation (3); the BFI is expressed as follows:

where BFI_P denotes the BFI of patient *P*. If the number of patients included in group *W* is denoted as N_W, the grouping compensation factor of group *W*, CF_W, can be computed as

On the basis of the grouping compensation factor, the MBFI of patient *P* can be updated by rewriting equation (3), as follows:

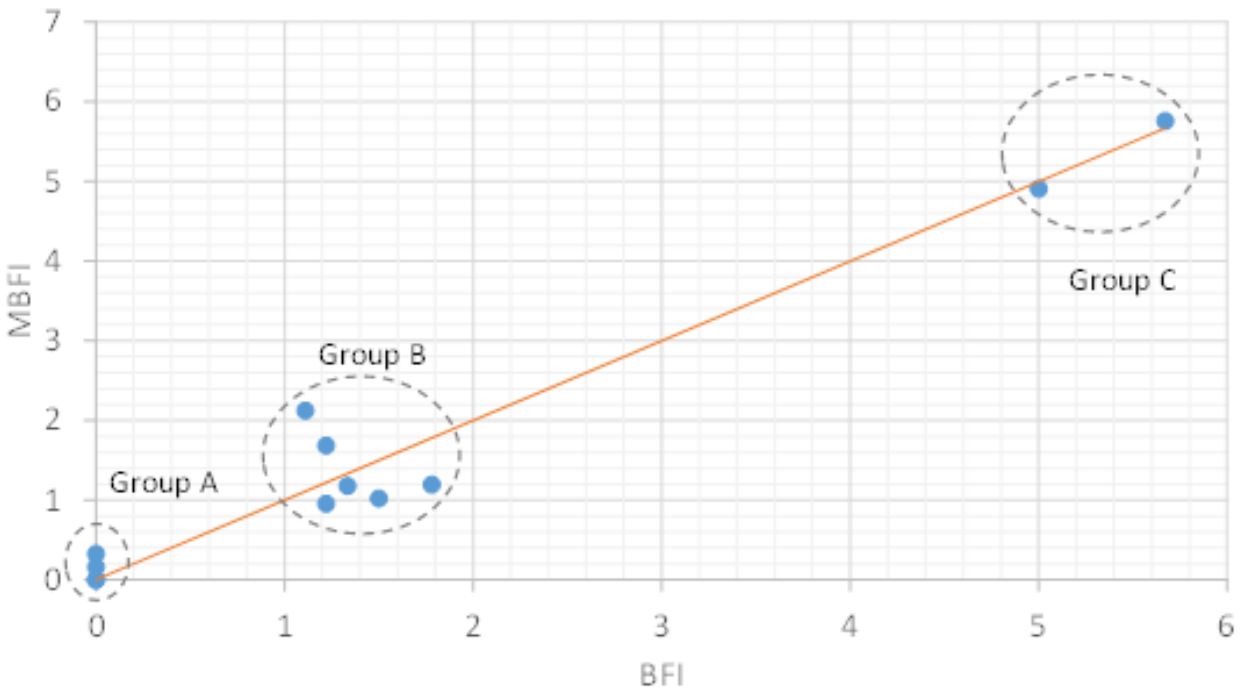
If the resultant MBFI was <0, the value of MBFI was fixed at 0.

Figure 6 shows the MBFI distributions for the BFIs of all patients. MBFI can be close to BFI, especially in groups A and C. For group B, the individual differences affect the approximation between MBFI and BFI in the presence of more patients. To evaluate the performance of MBFI, Table 3 presents the results of mapping errors in terms of MAE and RMSE. The MAE and RMSE are defined as follows:

As shown in Table 3, the MAE and RMSE values for group B were approximately 0.5, and the MAE and RMSE values for groups A and C were approximately 0.1. The MAE and RMSE values of all patients were 0.3 and 0.41, respectively. As MAE is the target mapping error metric, the mapping error rate of MBFI is (0.3/10)×100%=3%, where the BFI range is 0-10. According to the resultant MAE and RMSE, the MBFI provides a fair solution for estimating the cancer-related fatigue using PPG-based HRV parameters.

Figure 6. Mapped Brief Fatigue Inventory distribution. The red solid line represents mapped Brief Fatigue Inventory=Brief Fatigue Inventory. BFI: Brief Fatigue Inventory; MBFI: mapped Brief Fatigue Inventory.



XSL•FO
RenderX

Table 3. Mapping error statistics.

Metric	MAE[a]	RMSE[b]
Group A	0.11	0.18
Group B	0.49	0.56
Group C	0.09	0.09
Total	0.3	0.41

[a]MAE: mean absolute error.

[b]RMSE: root mean square error.

Discussion

Principal Findings

Analysis of the HRV data of the 12 patients with lung cancer enrolled in this study showed that the LF-HF ratio can be highly correlated with the subjective BFI, particularly that measured during sleeping time. On the basis of the correlation results, we further derived a BFI mapping equation based on the LF-HF ratio to estimate the target BFI. The analytical results showed that the total mapping error rate was 3% and that the LF-HF ratio can be considered a fair indicator to evaluate the degree of cancer-related fatigue during cancer treatment.

HRV pinpoints direct and indirect heart dynamics and respiration cycle alterations that couple differently during sleep and wakefulness across age and sex [26]. Environmental changes (eg, during the day and night) also influence HRV patterns [27]. Hence, it is not surprising that the emergence or existence of a disease and its biochemical progress affect HRV. A systematic review and meta-analysis of HRV in the prediction of cancer survival suggested that higher vagal nerve activity might predict prolonged survival [8]. HRV can be used to detect general systemic inflammation status. The Toon Health Study, which investigated 1728 nonsmoking Japanese individuals aged 30-79 years between 2009 and 2012, reported that HRV parameters are highly correlated with an elevation of C-reactive protein, a marker of inflammation [28]. HRV is probably under multicenter control, and the brain and heart are the primary control sources. Studies have provided preliminary evidence that brain damage [9,29] and heart damage [10,30] alter HRV patterns. Although the brain-heart axis mechanisms remain largely unknown, the communication between these two major organs probably involves the crosstalk of neuroimmunological chemicals [31] and induces a rise of norepinephrine in the brain that causes subsequent fatigue and sleep disturbance [16].

The current clinical measurement of HRV is based on stationary devices in a denoised environment (such as a physiological laboratory with good environmental noise isolation). Hence, only restricted activity information can be collected during the examination, which normally lasts for 5 minutes. This largely reduces the sensitivity and specificity of these hospital-based measurements. As a result, such devices are unlikely to provide an informative picture of the actual fatigue condition of patients. A comprehensive understanding of the continuous physiological changes is necessary to examine complex phenomena, such as fatigue, either for diagnostic or research purposes. In addition, a comprehensive collection of HRV information of patients with cancer enables early detection of fatigue symptoms, especially when patients fail to reveal their problems. Moreover, it helps to reduce the workload of medical staff, particularly nurses.

The main contributions of this study are the theory and application domains. In the theory domain, there are two phases in HRV data, namely the active phase and the sleep phase. Accordingly, three objective HRV parameters, namely the LF-HF disorder (LF-HF >1) ratio of the sleep phase, average LF-HF ratio of the sleep phase, and LF-HF disorder (LF-HF <1) ratio of the active phase, are related to subjective BFI results. Consequently, the MBFI can be calculated as an objective cancer-related fatigue indicator from the combination of three LF-HF-related parameters to approximate the subjective BFI.

Limitations

This study was mainly limited by three issues. First, the sample size was relatively small and thus the power. A more rigorous design would have been to recruit a second sample to determine whether the results are generalizable. Second, because of the small sample size, we did not adjust the variance of HRV caused by the different treatments. Third, sweating on wearing the device might have caused discomfort to some patients, and others might have been worried about damaging the device during work that involved washing; thus, some data might have been lost when the patients likely removed their PPG watch device at other instances besides showering.

Conclusions

The LF-HF ratio was highly correlated with the cancer-related fatigue. For decades, self-reported subjective assessment tools have been used to measure fatigue levels of the patients. Although the assessments may reveal the direct feelings of the patients, the exact physiological conditions could be missed owing to conservative expression and poor dysregulation of the patients. Beyond revealing fatigue levels objectively, continuous HRV recordings through the PPG watch device and the defined parameters LF-HF can outline the active phase and sleep phase in patients with lung cancer who undergo chemotherapy or target chemotherapy, allowing a deduction of their sleep patterns. Additional studies that examine the similarity and diversity of the LF-HF ratio in patients with different types of cancer are warranted.

XSL•FO
RenderX

Acknowledgments

Funding for this study was provided by the Ministry of Science and Technology of Taiwan (grant MOST 108-2622-E-038-003-CC3). The device was obtained from Gadgle Creative Technology. This manuscript was edited by Wallace Academic Editing.

Conflicts of Interest

None declared.

References

1. Bower JE. Cancer-related fatigue--mechanisms, risk factors, and treatments. Nat Rev Clin Oncol 2014 Oct;11(10):597-609 [FREE Full text] [doi: 10.1038/nrclinonc.2014.127] [Medline: 25113839]
2. Ma Y, He B, Jiang M, Yang Y, Wang C, Huang C, et al. Prevalence and risk factors of cancer-related fatigue: a systematic review and meta-analysis. Int J Nurs Stud 2020 Nov;111:103707. [doi: 10.1016/j.ijnurstu.2020.103707] [Medline: 32920423]
3. Piper BF, Cella D. Cancer-related fatigue: definitions and clinical subtypes. J Natl Compr Canc Netw 2010 Aug;8(8):958-966 [FREE Full text] [doi: 10.6004/jnccn.2010.0070] [Medline: 20870639]
4. Runowicz CD, Leach CR, Henry NL, Henry KS, Mackey HT, Cowens-Alvarado RL, et al. American cancer society/American society of clinical oncology breast cancer survivorship care guideline. J Clin Oncol 2016 Feb 20;34(6):611-635. [doi: 10.1200/JCO.2015.64.3809] [Medline: 26644543]
5. NCCN Clinical Practice Guidelines in Oncology: Cancer-related Fatigue. National Comprehensive Cancer Network. 2020. URL: https://www.nccn.org/professionals/physician_gls/pdf/fatigue.pdf [accessed 2021-06-11]
6. Kamen P. Heart rate variability. Aust Fam Physician 1996 Jul;25(7):1087-9, 1091. [Medline: 8768275]
7. Pierpont GL, Adabag S, Yannopoulos D. Pathophysiology of exercise heart rate recovery: a comprehensive analysis. Ann Noninvasive Electrocardiol 2013 Mar;18(2):107-117 [FREE Full text] [doi: 10.1111/anec.12061] [Medline: 23530480]
8. Zhou X, Ma Z, Zhang L, Zhou S, Wang J, Wang B, et al. Heart rate variability in the prediction of survival in patients with cancer: a systematic review and meta-analysis. J Psychosom Res 2016 Oct;89:20-25. [doi: 10.1016/j.jpsychores.2016.08.004] [Medline: 27663106]
9. Keren O, Yupatov S, Radai MM, Elad-Yarum R, Faraggi D, Abboud S, et al. Heart rate variability (HRV) of patients with traumatic brain injury (TBI) during the post-insult sub-acute period. Brain Inj 2005 Aug 10;19(8):605-611. [doi: 10.1080/02699050400024946] [Medline: 16175814]
10. Martínez-Navarro I, Sánchez-Gómez JM, Collado-Boira EJ, Hernando B, Panizo N, Hernando C. Cardiac damage biomarkers and heart rate variability following a 118-km mountain race: relationship with performance and recovery. J Sports Sci Med 2019 Dec;18(4):615-622 [FREE Full text] [Medline: 31827345]
11. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
12. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
13. Pereira T, Moreira T, Almeida PR, Cunha JP, Aguiar A. Fine Grained Stress Assessment in Ecological Conditions. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC'15; March 31, 2015; Milano, Italy URL: https://paginas.fe.up.pt/~voce/docs/EMBC2015.pdf
14. Ollson E. Heart Rate Variability in Stress-related Fatigue: Adolescent Anxiety and Depression and Its Connection to Lifestyle. Sweden: University of Uppsala. 2010. URL: http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A345553&dswid=-8365 [accessed 2020-06-14]
15. Assad S, Ding F, Fu N, Xu YJ. Correlating Heart Rate Variability with Mental Fatigue. Worchester Polytechnic Institute, Massachusetts. 2012. URL: https://digital.wpi.edu/downloads/t722hb42z [accessed 2021-06-11]
16. Fagundes CP, Murray DM, Hwang BS, Gouin JP, Thayer JF, Sollers JJ, et al. Sympathetic and parasympathetic activity in cancer-related fatigue: more evidence for a physiological substrate in cancer survivors. Psychoneuroendocrinology 2011 Sep;36(8):1137-1147 [FREE Full text] [doi: 10.1016/j.psyneuen.2011.02.005] [Medline: 21388744]
17. Vicente J, Laguna P, Bartra A, Bailón R. Drowsiness detection using heart rate variability. Med Biol Eng Comput 2016 Jun;54(6):927-937. [doi: 10.1007/s11517-015-1448-7] [Medline: 26780463]
18. Carskadon MA, Dement WC. Normal human sleep: an overview. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. Philadelphia, PA: Saunders; 1994.
19. Elsenbruch S, Harnish MJ, Orr WC. Heart rate variability during waking and sleep in healthy males and females. Sleep 1999 Dec 15;22(8):1067-1071. [doi: 10.1093/sleep/22.8.1067] [Medline: 10617167]
20. Busek P, Vanková J, Opavský J, Salinger J, Nevsímalová S. Spectral analysis of the heart rate variability in sleep. Physiol Res 2005;54(4):369-376 [FREE Full text] [Medline: 15588154]
21. Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. Sleep Med Rev 2012 Feb;16(1):47-66. [doi: 10.1016/j.smrv.2011.02.005] [Medline: 21658979]

XSL•FO
RenderX

22. Jarrin D, McGrath JJ, Giovanniello S, Poirier P, Lambert M. Measurement fidelity of heart rate variability signal processing: the devil is in the details. Int J Psychophysiol 2012 Oct;86(1):88-97 [FREE Full text] [doi: 10.1016/j.ijpsycho.2012.07.004] [Medline: 22820268]
23. Bánhalmi A, Borbás J, Fidrich M, Bilicki V, Gingl Z, Rudas L. Analysis of a pulse rate variability measurement using a smartphone camera. J Healthc Eng 2018;2018:4038034 [FREE Full text] [doi: 10.1155/2018/4038034] [Medline: 29666670]
24. Bolanos M, Nazeran H, Haltiwanger E. Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. Conf Proc IEEE Eng Med Biol Soc 2006;2006:4289-4294. [doi: 10.1109/IEMBS.2006.260607] [Medline: 17946618]
25. Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Goncalves L, et al. Can PPG be used for HRV analysis? Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:2945-2949. [doi: 10.1109/EMBC.2016.7591347] [Medline: 28268930]
26. Zorko A, Frühwirth M, Goswami N, Moser M, Levnajić Z. Heart rhythm analyzed via shapelets distinguishes sleep from awake. Front Physiol 2019 Jan 17;10:1554 [FREE Full text] [doi: 10.3389/fphys.2019.01554] [Medline: 32009972]
27. Nano M, Fonseca P, Overeem S, Vullings R, Aarts RM. Lying awake at night: cardiac autonomic activity in relation to sleep onset and maintenance. Front Neurosci 2019;13:1405 [FREE Full text] [doi: 10.3389/fnins.2019.01405] [Medline: 32009886]
28. Saito I, Hitsumoto S, Maruyama K, Eguchi E, Kato T, Okamoto A, et al. Impact of heart rate variability on C-reactive protein concentrations in Japanese adult nonsmokers: the Toon Health Study. Atherosclerosis 2016 Jan;244:79-85. [doi: 10.1016/j.atherosclerosis.2015.10.112] [Medline: 26595902]
29. Katz-Leurer M, Zohar N, Boum A, Keren O. Monitoring changes in heart rate, as an indicator of the cardiovascular autonomic nervous function, among patients at the sub-acute phase post-brain damage during a physiotherapy session: a preliminary investigation. Brain Inj 2014;28(1):127-131. [doi: 10.3109/02699052.2013.848381] [Medline: 24328738]
30. Meinardi MT, van Veldhuisen DJ, Gietema JA, Dolsma WV, Boomsma F, van den Berg MP, et al. Prospective evaluation of early cardiac damage induced by epirubicin-containing adjuvant chemotherapy and locoregional radiotherapy in breast cancer patients. J Clin Oncol 2001 May 15;19(10):2746-2753. [doi: 10.1200/JCO.2001.19.10.2746] [Medline: 11352968]
31. Ask TF, Lugo RG, Sütterlin S. The neuro-immuno-senescence integrative model (NISIM) on the negative association between parasympathetic activity and cellular senescence. Front Neurosci 2018;12:726 [FREE Full text] [doi: 10.3389/fnins.2018.00726] [Medline: 30369866]

Abbreviations

BFI: Brief Fatigue Inventory
ECG: electrocardiogram
HF: high frequency
HR: heart rate
HRV: heart rate variability
LF: low frequency
MAE: mean absolute error
MBFI: mapped Brief Fatigue Inventory
PNS: parasympathetic nervous system
PPG: photoplethysmography
REM: rapid eye movement
RMSE: root mean square error
SNS: sympathetic nervous system

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by S Shun, H Jim; comments to author 15.03.21; revised version received 20.03.21; accepted 04.05.21; published 05.07.21.

Please cite as:
Shih CH, Chou PC, Chou TL, Huang TW
Measurement of Cancer-Related Fatigue Based on Heart Rate Variability: Observational Study
J Med Internet Res 2021;23(7):e25791
URL: https://www.jmir.org/2021/7/e25791
doi:10.2196/25791
PMID:36260384

©Chi-Huang Shih, Pai-Chien Chou, Ting-Ling Chou, Tsai-Wei Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Experience Sampling and Programmed Intervention Method and System for Planning, Authoring, and Deploying Mobile Health Interventions: Design and Case Reports

Bruna Carolina Rodrigues Cunha[1], BSc, MSc, PhD; Kamila Rios Da Hora Rodrigues[2], BSc, MSc, PhD; Isabela Zaine[2], BSc, MSc, PhD; Elias Adriano Nogueira da Silva[2], BSc, MSc, PhD; Caio César Viel[3], BSc, MSc, PhD; Maria Da Graça Campos Pimentel[2], BSc, MSc, PhD

[1]Federal Institute of Education, Science and Technology of São Paulo, Capivari, Brazil

[2]Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, Brazil

[3]Sidia Institute of Science and Technology, Manaus, Brazil

Corresponding Author:
Bruna Carolina Rodrigues Cunha, BSc, MSc, PhD
Federal Institute of Education, Science and Technology of São Paulo
Avenida Doutor Ênio Pires de Camargo, 2971
Capivari, 13360-000
Brazil
Phone: 55 1921466700
Email: bruna.rodrigues@ifsp.edu.br

Abstract

Background: Health professionals initiating mobile health (mHealth) interventions may choose to adapt apps designed for other activities (eg, peer-to-peer communication) or to employ purpose-built apps specialized in the required intervention, or to exploit apps based on methods such as the experience sampling method (ESM). An alternative approach for professionals would be to create their own apps. While ESM-based methods offer important guidance, current systems do not expose their design at a level that promotes replicating, specializing, or extending their contributions. Thus, a twofold solution is required: a method that directs specialists in planning intervention programs themselves, and a model that guides specialists in adopting existing solutions and advises software developers on building new ones.

Objective: The main objectives of this study are to design the Experience Sampling and Programmed Intervention Method (ESPIM), formulated toward supporting specialists in deploying mHealth interventions, and the ESPIM model, which guides health specialists in adopting existing solutions and advises software developers on how to build new ones. Another goal is to conceive and implement a software platform allowing specialists to be users who actually plan, create, and deploy interventions (ESPIM system).

Methods: We conducted the design and evaluation of the ESPIM method and model alongside a software system comprising integrated web and mobile apps. A participatory design approach with stakeholders included early software prototype, predesign interviews with 12 health specialists, iterative design sustained by the software as an instance of the method's conceptual model, support to 8 real case studies, and postdesign interviews.

Results: The ESPIM comprises (1) a list of requirements for mHealth experience sampling and intervention-based methods and systems, (2) a 4-dimension planning framework, (3) a 7-step-based process, and (4) an ontology-based conceptual model. The ESPIM system encompasses web and mobile apps. Eight long-term case studies, involving professionals in psychology, gerontology, computer science, speech therapy, and occupational therapy, show that the method allowed specialists to be actual users who plan, create, and deploy interventions via the associated system. Specialists' target users were parents of children diagnosed with autism spectrum disorder, older persons, graduate and undergraduate students, children (age 8-12), and caregivers of older persons. The specialists reported being able to create and conduct their own studies without modifying their original design. A qualitative evaluation of the ontology-based conceptual model showed its compliance to the functional requirements elicited.

Conclusions: The ESPIM method succeeds in supporting specialists in planning, authoring, and deploying mobile-based intervention programs when employed via a software system designed and implemented according to its conceptual model. The

XSL•FO
RenderX

ESPIM ontology–based conceptual model exposes the design of systems involving active or passive sampling interventions. Such exposure supports the evaluation, implementation, adaptation, or extension of new or existing systems.

(J Med Internet Res 2021;23(7):e24278) doi:10.2196/24278

KEYWORDS

mobile apps; mHealth; intervention; experience sampling; method; monitoring; Experience Sampling and Programmed Intervention Method; experience sampling method; ecological momentary assessment; just-in-time adaptive intervention

Introduction

Many factors impact the adoption of mobile health (mHealth) tools by professionals and their target population, as observed in Australia [1], Canada [2], USA [3,4], and in Europe [5,6]. As an example, clinicians' concerns when considering an mHealth tool include usefulness, ease of use, compatibility, technical issues, content, personalization, convenience, strict data privacy, workload, workflow, communication, management support, and policies [7-9]. Such themes align with those highlighted by Chinese public hospitals' managers [10], including perceived ease of use, system security and reliability, top management support, and government policy.

Toward employing mHealth interventions, professionals can generally choose among 3 options: using apps designed for other activities such as peer-to-peer communication, using purpose-built apps specialized in the required intervention, or using apps based on methods such as the experience sampling method (ESM) [11-13], and its descendent ecological momentary assessment (EMA) [14], including those exploring just-in-time adaptive interventions (JITAIs) [15]. The first alternative allows professionals to adapt their protocols to take advantage of popular apps [16] and to employ conventional SMS text messaging usually available to the underprivileged [17-19]. However, because interventions may require sending or collecting multiple types of questions and media and demand careful planning [20], deploying nonspecialized apps demands both adaptations in the protocol and overcoming obstacles when monitoring progress.

The second alternative led to the design of a wide range of mHealth-specialized apps [21-24] that enable reproducing interventions accurately. Their design engenders a dependency relationship between specialists and software developers. Moreover, specialized apps have little potential for reuse.

The third alternative involves using apps based on methods such as the ESM and EMA, as in the works surveyed by van Berkel et al [25]. Examples include studies [26-30] that employed the LifeData [31], the movisensXS [32], or the Mobile EMA [33] systems based on data collection methods. Additionally, ecological momentary interventions (EMIs) or JITAIs support interventions involving contextual data used for personalization according to users' needs [15,34,35].

In a complementary approach, if professionals were able to create their own apps [36], they could focus on the methodological processes of their work. While the ESM methods offer important guidance, current systems [31-33] do not expose their design at a level that promotes replicating, specializing, or extending their contributions as demanded in many areas. Thus, a twofold solution is required: a method that directs specialists in planning an mHealth intervention program themselves, and a model that guides specialists in adopting existing solutions while advising software developers on building new ones.

The 2 main objectives of this study are to design the Experience Sampling and Programmed Intervention Method (ESPIM), formulated toward supporting specialists in deploying mHealth interventions, and the ESPIM model, which guides specialists in adopting existing solutions and advises software developers on how to build new ones. A subsidiary goal is to conceive and implement a software platform allowing specialists to be users who plan, create, and deploy interventions (ESPIM system).

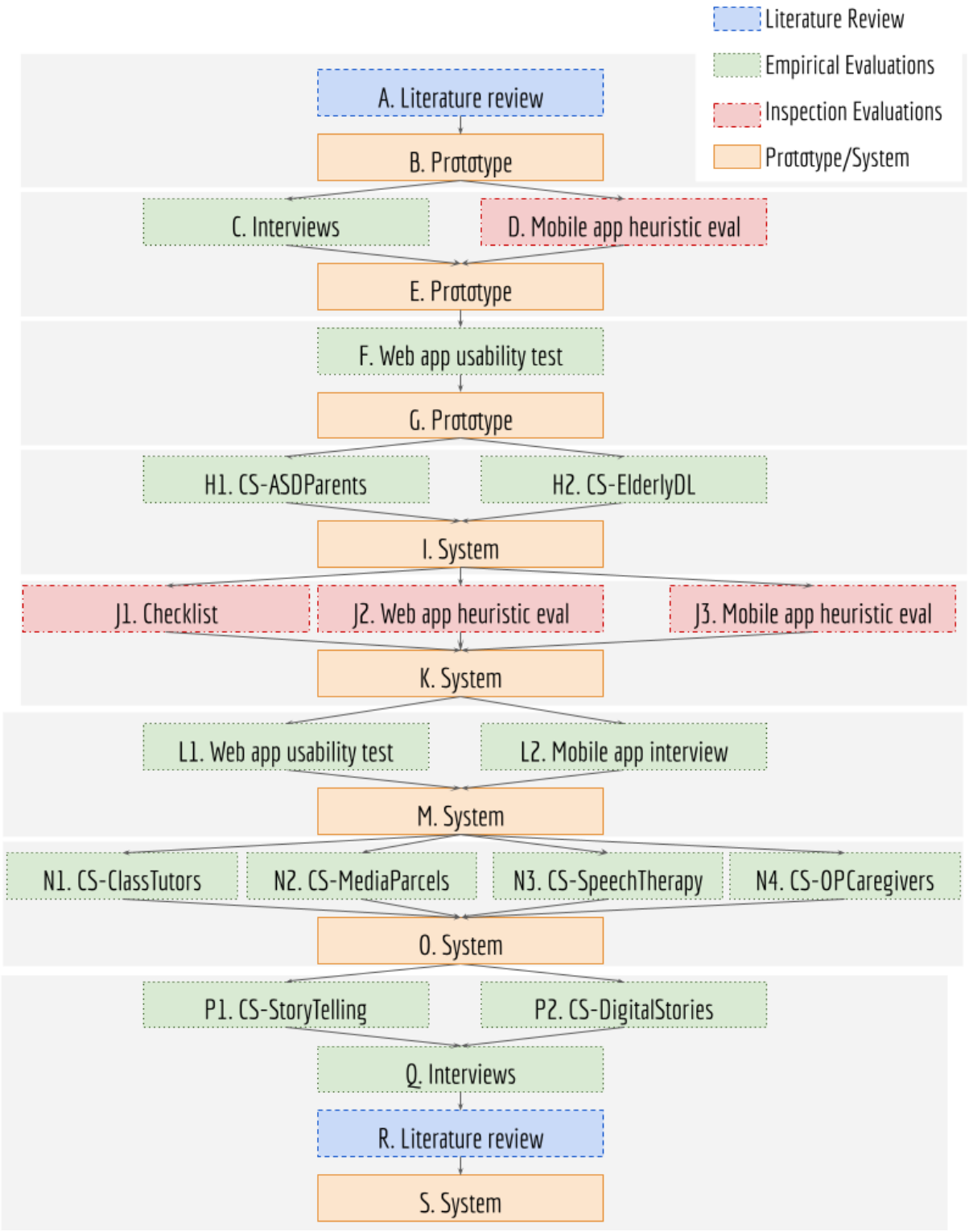
Methods

Overview

For designing the mobile-based ESPIM method, we adopted an iterative approach of co-design considering the participatory design practices [37,38]. Besides continuous review of state-of-the-art literature, the procedures adopted encompass early software prototype and predesign interviews with health specialists, iterative design sustained by a software system, support to real case studies, and postdesign interviews (Figure 1). The evaluation of ESPIM software employed the heuristic evaluation method [39] (see Textbox SM1 in Multimedia Appendix 1), usability tests [40] (see Textbox SM2 in Multimedia Appendix 1), the Semantic Differential Scale [41] (see Textbox SM3 in Multimedia Appendix 1), and the User Experience Questionnaire (UEQ) [42] (see Textbox SM4 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Study Methods and Workflow. CS: case study.



The methods used in this study were approved by the Brazilian Research Ethics Committee under case number 57875016.3.0000.5390. Study participation was voluntary and respected anonymity.

Recruitment

Users were involved in the participatory design as specialists or target users. In the case of specialists, they form a convenience sample recruited via email at nearby research departments. Specialists took part as intervention planners; inclusion criteria were being a health or education professional and experience in planning and delivering interventions. Education specialists participated as educational issues are part of eHealth [43] and corresponding interventions influence, among others, socialization, cognitive support, and mental health [44-46].

XSL•FO
RenderX

Regarding the target users who participated in the case studies, specialists handled their recruitment as they were participants in their interventions.

The study also had the contribution of specialists in human–computer interaction (HCI). They were recruited via email from nearby research departments and software companies. Inclusion criteria were background in HCI and experience in conducting heuristic evaluations.

Professional designers, specialists in user interface (UI) and user experience (UX), were also recruited to design refined interfaces and validate them after the implementation. They were recruited via email from nearby software companies.

Participatory Design

Evolution of ESPIM

Following a participatory design approach, the ESPIM method evolved according to requirements gathered from literature reviews, interviews, prototyping, and implementation of software instances of the associated model, and gradual usability evaluations (Figure 1). As observed by Byambasuren et al [47], usability is a main barrier for prescription and adoption of mHealth apps, especially for older people. To overcome these barriers, usability and UX were evaluated through empirical and inspection-based evaluations, which provided new requirements and detailed existing ones [48-51].

Interview (Predesign)

During requirement collection, specialists of different domains participated in semistructured interviews (Figure 1C and Multimedia Appendix 2) to discuss their needs in carrying out remote data collection and interventions, to answer a survey of the difficulties they faced in data collection and interventions, to present the system prototype as a potential solution, and to collect requirements. These professionals worked, among others, with children and adults with typical and atypical development, older people, pregnant women, and individuals with motor impairments. All participants conducted academic research in their areas. They were individually interviewed at their workplaces (by IZ and KRHR). Each meeting lasted 1 hour and 30 minutes on average. Interviews were performed with all participants who accepted the invitation.

First Heuristic Evaluation (App Prototype)

Four HCI specialists (Figure 1D), 2 researching usability for older people and 2 accessibility, conducted the first heuristic evaluation of the mobile interface of the ESPIM app prototype (question based at the time) [49]. The session lasted 24 hours so specialists could evaluate the 4 daily temporal triggers besides initiating the program themselves. The questions were related to daily routines and aimed to connect the trigger time with locations, daily activities, information technologies used, and ongoing activities.

First Usability Test (Web Application Prototype)

The first usability test performed on the web interface of ESPIM prototype (Figure 1F) aimed at answering: Do the specialists clearly understand what is the system and its purpose? Do the specialists face difficulties creating intervention programs using the system? Are the specialists able to complete all stages of creating intervention programs, including the setup phase? The test protocol (see Textbox SM5 in Multimedia Appendix 1) evaluated the understanding and the performance of the prototype from the point of view of health specialists using the metrics task execution time, number of steps to complete tasks, number of completed tasks, number of errors, and overall satisfaction.

We provided a hypothetical scenario *Geriatrician* outlining the tasks conducted by a specialist (geriatrician) who plans an intervention program to an older person with cognitive impairment who is attended by a caregiver (see Table SM1 in Multimedia Appendix 1). The scenario comprises 8 tasks requested of the older person, who is aided by the caregiver if needed. We specified the most complex task as a diagram (see Figure SM1 in Multimedia Appendix 1).

Checklist

The UI/UX designers who designed graphical interfaces executed a checklist-based evaluation (Figure 1J). Upon interacting with the system, they enlisted improvements.

Heuristic Evaluation

HCI specialists performed an inspection-based evaluation on the web application interface (Figure 1L1). The evaluators received an email with instructions, a task checklist, links, and files. The links led to the informed consent form, a profile survey, and to the ESPIM web application. The files contained instructions on how to conduct the evaluation, a template for reporting issues identified along with the corresponding heuristics, and the hypothetical scenario *Geriatrician*, comprising the tasks to be analyzed (see Table SM1 in Multimedia Appendix 1). After individual inspections, the evaluators met to discuss the problems found in the interface and produced a consolidation report.

The second heuristic evaluation of the mobile app followed the same protocol, using the same files (Figure 1J3). Additionally, the evaluators received the app installation file which included an intervention program and corresponding task checklist of the hypothetical scenario *Nutrition*. Four evaluators inspected the app performing 6 tasks: install, initiate and give the permissions requested by the ESPIM app, log in with your Google account, start the “Nutritional Data Collection” intervention program and navigate through all screens planned by the nutritionist, explore the app settings, and disconnect from the app.

Second Usability Test (Web Application) and Interview (App)

The second usability test on the web application (Figure 1L1) and the interview about usability and interaction aspects of the mobile app (Figure 1L2) were conducted at the same time. The questions and the protocol were those used in the first usability test.

For testing the web application, we provided the scenario “Monitoring and evaluating the performance of the older people in digital literacy courses through remotely programmed interventions” (see Table SM2 and Figure SM2 in Multimedia

XSL•FO
RenderX

Appendix 1). Aspects of the web interface elements evaluated were ease of use, memorization, easiness to "undo" actions, learnability during use, intuitiveness, feedback/error messages, information organization, arrangement of interface elements, available features, and interface design. Participating specialists answered a 7-point Likert scale (1="Awful" to 2="Excellent") for each aspect. The specialists also provided a self-evaluation of their performance using the system by responding to 7 affirmative sentences using a 5-point Likert scale (1="Strongly agree" to 5="Strongly disagree"): I easily completed the required tasks, I completed the tasks rapidly and efficiently, I would need someone's support to use the system, I felt more productive during the interaction with the system, I easily found the information and functionalities that I needed, I needed to thoroughly think or remember before completing the tasks, and I would recommend the system to other people.

The specialists had previous experience as instructors or tutors in mobile digital literacy courses for elderly individuals (Figure 1H2) as part of the Case Study ElderlyDL (Table 1). They were interviewed regarding their learning while observing older people interacting with the app and commonly reported complaints. The questions asked were "Which errors or infrastructure problems were found? Which devices did not work? Which frustrations were observed while older people interacted with the app? For each type of task: What were the main difficulties of the older people? What were the issues? What could be better? What is good and should not change?."

Table 1. Eight case studies.

Case study	Specialists involved	Participants	Short description
CS-ASDParents	2 psychologists; 1 computer scientist	3 families (3 children with autism spectrum disorder and 3 parents)	Promoting engagement in educational activities between children diagnosed with autism spectrum disorder and their parents.
CS-ElderlyDL	3 gerontologists, 2 computer scientists, 1 psychologist, 1 statistician	365 older people (age 60+)	Supporting mobile digital literacy courses for elderly.
CS-ClassTutors	1 gerontologist, 1 computer scientist	12 tutors (graduates/undergraduate students)	Analyzing elderly digital literacy courses using tutors' feedback.
CS-MediaParcels	1 psychologist, 1 computer scientist	1 family (1 father and 2 children); 3 elderly friends (age 60+)	Encouraging multimedia interventions to promote social connection among elderly.
CS-SpeechTherapy	1 speech therapist	5 children (age 8-12)	Deploying speech therapy homework for children.
CS-OPCaregivers	1 occupational therapist	30 caregivers of older people	Providing informative contents for caregivers of elderly with dementia.
CS-OPStorytelling	2 computer scientists, 1 gerontologist	15 older people (age 60+)	Enabling the creation of digital storytelling by seniors.
CS-StoryReading	1 psychologist	45 children (age about 10)	Developing digital stories for children with reading disabilities.

Case Studies

Overview and Approval

The ESPIM method evolved supporting real case studies conducted by specialists (Figure 1) who used the method via the associated ESPIM system to manage interventions with their populations of interest [52,53]. These case studies are part of the empirical evaluations of the ESPIM system (Table 1).

Each case study was submitted and approved by the Brazilian ethics committee and all data were anonymized. Their common study protocol included the following: signing an informed consent form, filling out a pretest profile survey, filling out a posttest questionnaire about the interaction experience, and participating in a semistructured interview at the end of the study. Each specialist applied specific evaluation forms of their respective fields to analyze the results of their studies.

Case Study ASDParents

The case study ASDParents studied engagement in educational activities between children with autism spectrum disorder and their parents (Figure 1H1). Parents were instructed to conduct at least one out of three planned educational activities at home with their children, once a day, during the 6 weeks: the first 3 weeks used conventional paper-based written instructions, whereas the last 3 employed ESPIM. Psychologists employed ESPIM to send text and video tasks and to monitor task accomplishment and performance. All children studied at a Brazilian nongovernmental organization with a 2-hour/week workload. Three families (3 children and 3 parents) participated [54].

Case Study ElderlyDL

The case study ElderlyDL offered mobile digital literacy courses to older people (Figure 1H2). Computer science, psychology, and gerontology specialists employed the web application to design intervention programs as homework for the older people, who received and responded to the tasks via an app. During 13 weeks, the app sent, on weekdays, a notification around 7 pm alerting about the homework. The study involved 365 older people [48,50,55].

Case Study ClassTutors

Case study ClassTutors collected feedback from tutors assisting 3 instructors providing digital literacy courses for older people (Figure 1N1). The tutors were undergraduate and graduate students. One instructor (gerontologist) used the web application to guide tutors in evaluating the effectiveness of the classes and

XSL•FO
RenderX

identifying situations of stress or struggle. The app sent 1 notification asking for feedback after the weekly class, the intervention being available throughout the week. A total of 12 tutors participated in this study for 4 months.

Case Study MediaParcels

In the Case study MediaParcels (Figure 1N2), 1 psychologist employed ESPIM as a multimedia exchange tool to investigate the impact of social interventions among older people and their connections. The specialist designed interventions to encourage participants to share self-revelations and media with affective content. The psychologist requested content from 1 participant, annotated the content with the meaning embedded in the original request, and forwarded it to the participant's connections. One study involved family members (father and 2 children) and another study involved 3 elderly friends. Both studies lasted 2 weeks [56].

Case Study SpeechTherapy

The case study SpeechTherapy was applied in the clinical context of speech therapy (Figure 1N3). The specialists planned reading, writing, and comprehension tasks to complement activities conducted at the clinic. They applied remote interventions with 5 patients (aged 8-12 years) with reading or writing issues. Five children participated in the 5-month case study.

Case Study OPCaregivers

The case study OPCaregivers delivered information to caregivers of older persons with dementia toward guiding and qualifying the care provided (Figure 1N4). One occupational therapist created interventions to present information related to feeding, personal hygiene, guidelines for maintaining a structured and stimulating routine, and tips for managing behavioral symptoms. Thirty caregivers participated in this study.

Case Study OPStorytelling

The case study OPStorytelling employed ESPIM interventions to guide the creation of video stories by users with little experience in producing digital content, especially older people (Figure 1P1). One specialist used ESPIM to create interventions as "storytelling scripts" that combined requests for text, video, image, and audio assets. A dedicated service, integrated into the ESPIM software, received the media assets, generated the corresponding video, and uploaded it to a YouTube private channel. This study employed 2 workshops to teach 15 older people to produce video-based narratives [57].

Case Study StoryReading

In the case study StoryReading, a specialist in psychology used the ESPIM system as a tool to create instructional programs in the form of text-based stories, augmented with images and animations (Figure P21). The target users were children with reading difficulties. The goal was to improve reading comprehension by delivering stories integrated with questions and corresponding interactive feedback.

Interview (Postdesign)

A member of the ESPIM team (BCRC) interviewed the specialists after their studies (Figure 1R). The specialists responded to a semistructured 2-part interview: area of expertise and related studies, and how they modeled interventions in their studies. The latter aimed to elicit how specialists designed, delivered, and monitored interventions using the ESPIM system.

Results

User Statistics

Predesign interviews (Figure 1C) involved 12 health specialists: 5 psychologists (4 specialized in special education and 1 in behavioral psychotherapy), 3 nurses (1 specialized in public health, 1 in health sciences, and 1 in mental health), 2 physicians (1 specialized in obstetrics and gynecology and 1 in psychiatry, geronto-psychiatry, and neurology), 1 physiotherapist (observer in neuropediatrics), and 1 occupational therapist (specialized in public health).

The first heuristic evaluation of the mobile app (Figure 1D) was conducted by 4 HCI specialists (2 specialized in usability for the older people and 2 in accessibility).

The first web application usability test (Figure 1F) involved 5 psychologists (3 specialized in special education; 1 in science, technology, and society; and 1 in biology).

The checklist-based evaluation was realized by 2 UI/UX designers (Figure 1J1).

The heuristic evaluation of the web application (Figure 1J2) and the second heuristic evaluation of the mobile app (Figure 1J3) involved 4 HCI specialists: 1 inexperienced (never performed this kind of evaluation), 1 had intermediary experience (conducted 3 evaluations), and 2 were experienced (executed more than 3 evaluations).

The second usability test of the web application (Figure 1L1) and the interview about usability aspects of the app (Figure 1L2) were performed, in the same session, with 5 gerontologists and 1 occupational therapist.

Postdesign interviews (Figure 1Q) involved 8 specialists: 2 psychologists, 2 gerontologists, 1 occupational therapist, 1 speech therapist, and 2 computer scientists who offered digital literacy courses for older persons in collaboration with one of the gerontologists.

Target users who took part in the interventions via the ESPIM mobile app included 431 older persons, 30 caregivers of older persons, 5 children, 12 undergraduate/graduate students, and 3 families with children with autism spectrum disorder (Table 1).

Participatory Design

Overview

The ESPIM comprises (1) a list of requirements for mHealth experience sampling and intervention-based methods and systems, (2) a 4-dimension planning framework, (3) a 7-step-based process for planning interventions, and (4) an ontology-based conceptual model. The ESPIM system and the cases study reports complement the contribution.

XSL•FO
RenderX

ESPIM's Functional Requirements

The functions demanded from ESPIM were elicited using literature review, predesign interviews, iterative prototyping, and postdesign interviews. The resulting functional requirements (Table 2) concern creating and managing (1) the "intervention programs," (2) the "persons" involved as observers and participants, (3) the "events" constituting the program and comprising triggers and tasks, (4) the "active tasks" specialists request to target users, (5) the use of "sensors" to capture data or trigger tasks, or both, and (6) the access to the "results."

XSL•FO
RenderX

Table 2. Functional requirements (FRs) for ESPIM.

FR	Description
Intervention program	
FR01	Enable creation, management, and reuse of intervention programs.
FR02	Enable definition of open or fixed beginning and ending dates for programs.
FR03	Enable organization of programs into phases composed by different events.
Person	
FR04	Enable registration and management of observers.
FR05	Enable registration and management of participants.
FR06	Enable collaborative management of programs.
FR07	Provide user authentication/authorization with roles and permissions.
FR08	Enable multiple associations among participants and programs.
FR09	Enable creation of contact lists related to privacy control.
FR10	Enable the association of relationships among participants (eg, communication).
FR11	Enable importing participants' data from external sources.
Event	
FR12	Enable creation of active tasks and sensor-based sampling.
FR13	Enable the association of triggers to events.
FR14	Enable configuration of intrusiveness level in triggers.
FR15	Provide trigger types: self- and specialist-initiated, temporal, contextual, random.
FR16	Enable definition of triggers' timeout.
FR17	Enable color-coding events.
FR18	Enable annotations and follow up via participant's app interface.
FR19	Enable configuration of triggers disabling when a condition is fulfilled.
FR20	Enable configuration of alert to inform when a participant did not answer a trigger.
FR21	Enable configuration of alert to inform when a participant answered a trigger.
FR22	Enable configuration of triggers rescheduling when a condition is satisfied.
FR23	Enable configuration of alert to inform of a specific answer (eg, risky behavior).
FR24	Enable configuration of automatic processing of responses (eg, condition based).
Active tasks	
FR25	Provide active tasks type message, question, media request, and external app launch.
FR26	Enable active tasks containing multimedia stimuli and emphases-enriched text.
FR27	Provide open-ended, multiple/single-choice (including pictures as choices) questions and scales (eg, Likert, sorting scale, grid).
FR28	Enable definition of mandatory active tasks.
FR29	Enable configuration of interaction flows (skip, branch, and loop).
Sensor-based sampling	
FR30	Enable configuration of sensor-based sampling while interacting with active tasks.
FR31	Enable definition of time intervals for sensor-based sampling.
FR32	Enable configuration of automated sensor-based sampling.
FR33	Enable sampling from software- or hardware-based sensors (eg, mobile and wearable devices, accessories, home sensors).
Results	

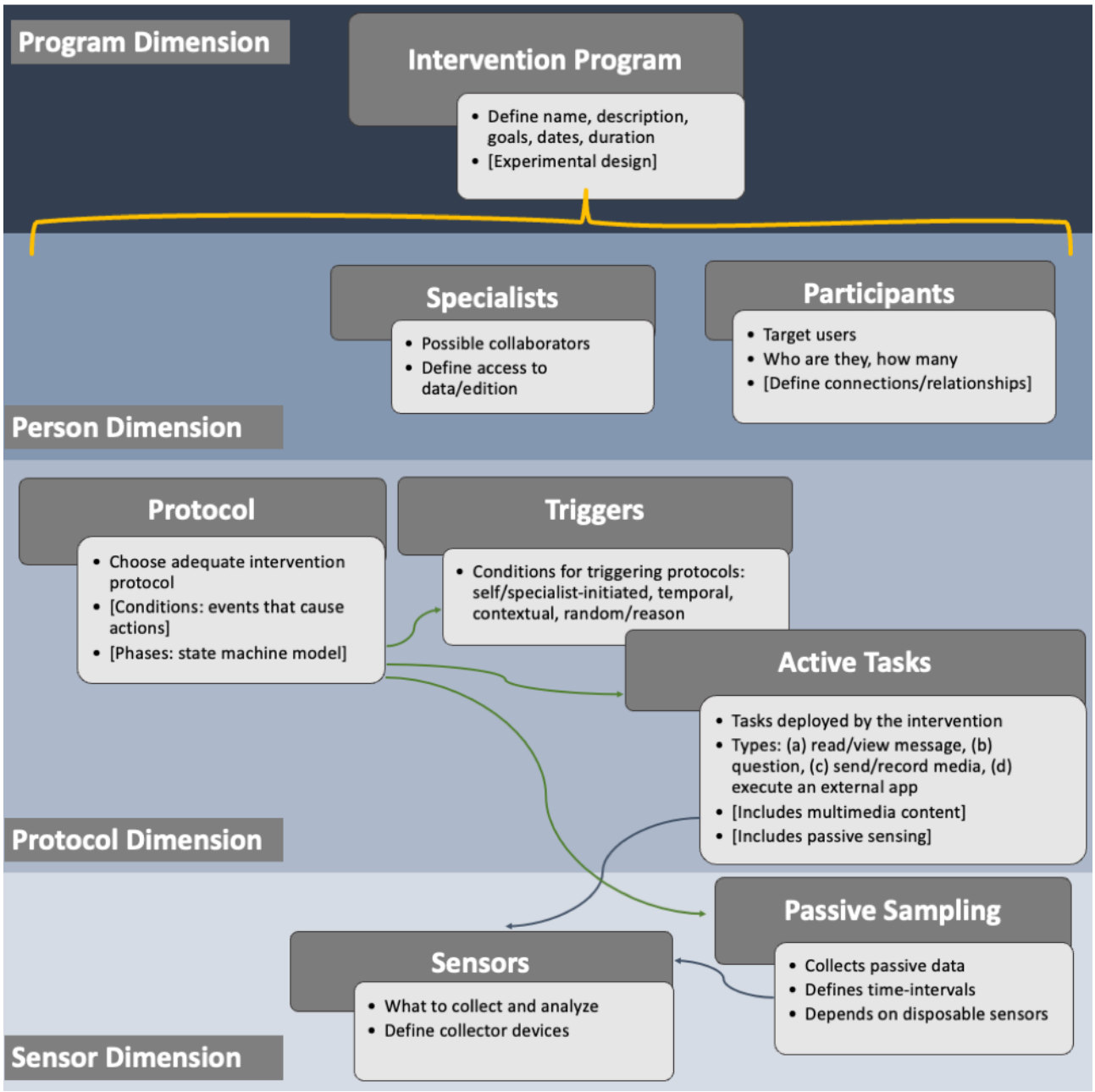
XSL•FO
RenderX

FR	Description
FR34	Provide different filter modes for results' visualization.
FR35	Provide date, time, and duration of the corresponding result.
FR36	Enable download of the results.
FR37	Enable configuration of automatic analysis of responses by dedicated algorithms.

ESPIM 4-Dimension Planning Framework

Defined upon the requirements, an ESPIM intervention program involves 4 dimensions: program, person, event, and sensor (Figure 2). The *program* dimension comprises the intervention program as a whole, and characterizes general intervention settings, such as name, definition, description, goals, duration, and, in research-based situations, its experimental design.

Figure 2. ESPIM Intervention dimensions: Program, Person, Event and Sensor.



The *person* dimension comprises the specialists in charge of the program and their target users, also called participants. Both professionals and researchers may work with collaborators who require distinct access control levels. Moreover, establishing target users is a key aspect of an intervention. In analytical work, the number of participants and their characterization are crucial, while individualized interventions and use case scenarios may demand the participation of other individuals (eg, parents, caretakers, partners). Relationships between participants and

XSL•FO
RenderX

these connections, along with delimited data sharing, might be considered in the program.

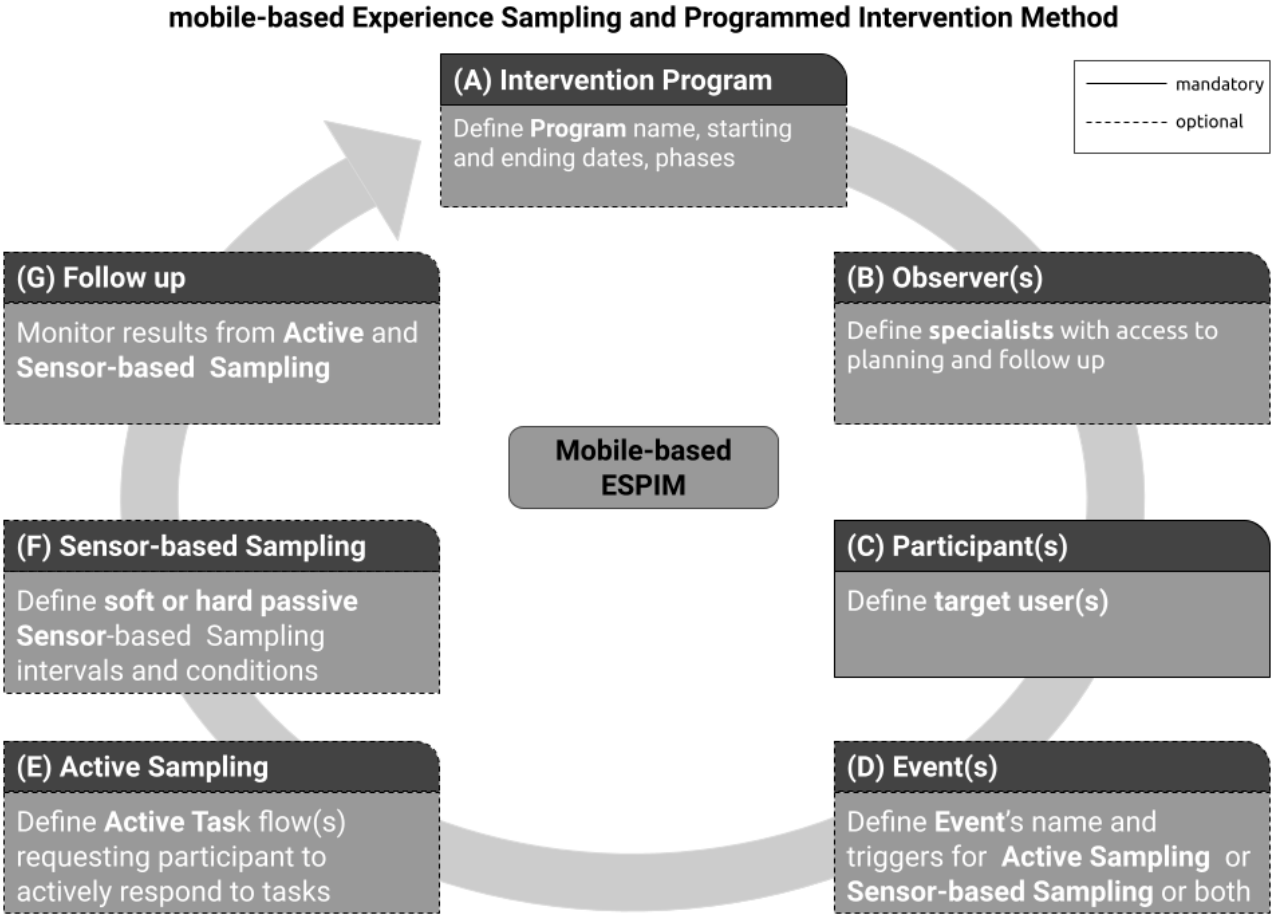
The *protocol* dimension comprises features provided by mobile technologies, aggregating the flow(s) that constitute the intervention along with the corresponding time- or sensor-based triggers. Finally, the *sensor* dimension allows expressing sensor-based support both for data collection and for triggering events.

ESPIM 7-Step-Based Process

Overview

ESPIM directs specialists in iteratively planning a mobile-based intervention program (Figure 3). The 7-step-based process suggests first identifying the intervention program by a name and defining its overall duration (Figure 3A), followed by the identification of who will have access to the planning procedures (Figure 3B) and who the target users are (Figure 3C). The intervention program is then defined in terms of 1 or more events (Figure 3D) which combine triggers to sampling procedures formulated as active tasks (Figure 3E) or sensor-based sampling (Figure 3F). Once the intervention program is deployed, specialists monitor participant's interaction and data collected (Figure 3G). These procedures are detailed next along with the corresponding requirements (Table 2).

Figure 3. ESPIM 7-step process.



Intervention Program: Name and Duration

The first step (Figure 3A) is to identify the intervention program by a name so that specialists can refer to the program for deployment, management, reuse, and change (FR01 in Table 2).

At the time of design, the intervention program deployment dates may not be determined. A program may be designed to be reused at distinct times, with different participants. Thus, when the intervention program is devised, the definition of starting or ending dates is optional (FR02).

An intervention program may be designed and structured to be applied in separate stages or phases delimited by time or other completion conditions (FR03).

Observers: Persons Who Have Access to the Planning Procedures

The second step is to register other specialists with access to the program (FR04 and FR06). This enables cases in which an intervention program is collaboratively designed or deployed, or both, by more than 1 specialist. Specialists that are responsible to manage intervention programs are called observers, and the optional step "observers" (Figure 3B) is employed when collaborators other than the one creating the

XSL•FO
RenderX

intervention program should have access to its specification and results.

Participants: Persons Who Receive the Intervention Tasks via Their Mobile Device

The third step (Figure 3C) is to register the target users (participants) of the program (FR05). However, the actual target users may not have been determined when the intervention program is initially designed. Moreover, an intervention program may be designed to be reused with different participants at distinct times. Thus, when first creating the intervention program, the observer may not include real target users. Clearly, defining a participant is mandatory for deploying an intervention program. One strategy adopted by specialists in the case studies we report was to include themselves as participants, allowing them to test the intervention program before deployment by assuming a participant role (FR07).

Observers may include 1 or more target users in 1 or more programs (FR08). This characteristic is essential to provide flexibility to personalize programs for individualized monitoring or to monitor groups of target users. Specialists register participants' contact information and aliases targeting users' privacy (FR09).

Another type of participant represents persons having relationships to target participants, such as family members or caretakers, with whom specialists may interact. Thus, a requirement is support for relationship among participants and associating different events with distinct participant roles in the relationship, which should be personalized according to each case (FR10).

Finally, specialists may use existing information systems to import participants' data valuable to intervention planning. Therefore, a service-based approach should be considered for third-party system integration and information exchange (FR11).

Event(s): What Sets of Tasks Are Triggered, and When, in the Participant's Smartphone

An intervention program consists of 1 or more intervention events executed along a time frame. Each event (Figure 3D) is identified by a name which is used for reference within the intervention program itself or in other programs, warranting reuse of events in particular, additionally to the reuse of programs (FR01).

To each event, the observer associates a set of tasks to be put into effect in the participant's mobile device. The actual tasks are defined by means of an active set (Figure 3E) or via sensor-based sampling (Figure 3F), or both (FR12). In the first case, participants respond explicitly to a task, for instance, by answering a question or capturing a video. In the second case, data are gathered passively from the participant's smartphone, for instance, via sensors or automated logging routines.

Further, to each event the observer associates 1 or more triggers specifying the times in which event's tasks are to be executed in the participant's mobile device (FR13). This implies that, at the times planned by the specialists, the participant's smartphone receives a notification corresponding to the event. If the notification triggers a flow of active tasks (ie, intervention flow), a notification is presented in the participant's device at the level of intrusiveness specified (FR14); when the participant responds to the notification, the mobile app allows the participant to interact with the intervention flow. In cases in which the notification triggers a sensor-based sampling routine, the corresponding sensor or automated routine is executed.

Triggers should be of diverse types besides being time based (FR15) and should be associated with a timeout (FR16). Specialist-initiated triggers allow observers to launch tasks on their own initiative. Self-initiated triggers are necessary when specialists opt to allow users to execute an event at their own initiative by starting the mobile app themselves at any time. A random trigger is appropriated when specialists demand the event to start at unconventional times without a predetermined pattern. A contextual trigger is set off when a particular situation occurs, such as one defined by rules involving 1 or more conditions associated with physical sensors (eg, global positioning system coordinates or heart rate monitor) or software-based data (eg, agenda). In any case, the specialist may indicate that different events are of distinct types by using coding such as different colors (FR17). Moreover, an event that has been already completed may be indicated as such in the mobile app, along with other information that allows participants to be aware of their status in the intervention (FR18).

Furthermore, how a participant reacts to a trigger may be an important aspect to some studies; for instance, a trigger may need to be rescheduled or the observers should receive an alert when a participant did not answer (FR19-FR24). Contextual triggers (FR15) and automatic responses processing (FR24) are essential requirements if specialists consider ecological assessment for adaptive interventions (EMA and JITAI).

Intervention(s): What Flow of Tasks Are Explicitly Demanded via Participants' Devices

An intervention flow (ie, active set; Figure 3E) is specified by the observer (FR12) to be presented to the participant via a mobile app in a customizable flow of active tasks (FR25). An active task is an intervention-based component that may contain 1 or more multimedia stimuli allowing, among others, sending instructions and requesting information; text-based stimuli should support emphasis including bold and italics (FR26). A question is a type of active task which comprises different formats (FR27) and they may be mandatory or not (FR28). One type of question, single-choice question, allows associating a different flow with each of the alternatives defined, as a result, conditional parallel flows (FR29). An active task may interact with a third-party mobile app, for both activating that app with customized configuration and collecting data resulting from its execution (FR25). An active set may also trigger a passive collection of sensor data without the need for an explicit user intervention (eg, capturing the face expression during the task; FR30).

Passive Sampling: What Information Will Be Collected Without Explicitly Asking the User

Passive sampling is needed when the design of the intervention program makes use of passive data collection without interrupting the participant (Figure 3F). This can be initiated via the configuration of temporal intervals (FR31) or of

XSL•FO
RenderX

automated sensor or software-based data collection (FR32), or via association with an active set (FR30). The collection may use sensors and software executing in devices other than the participant's smartphone such as wearable devices, accessories, and home sensors (FR33).

Passive sampling may be executed without an associated active set, as illustrated by Harari et al [58] when collecting data from sensors and logs. When this is the case, the observer specifies the sensors to be used as well as the conditions and the intervals of the collection.

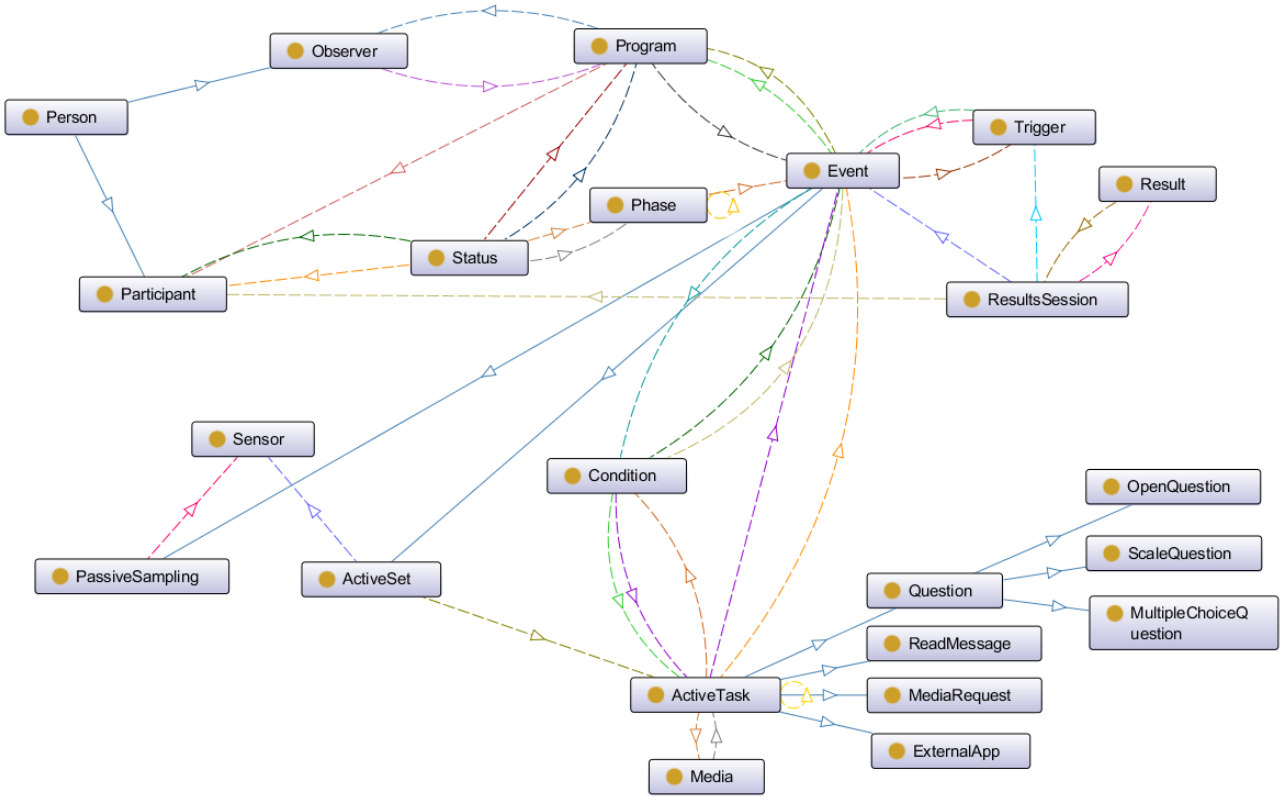
Results: When and How Participants Participated in the Intervention Program?

In some interventions, specialists demand monitoring how participants engage in the program to measure its impact or to adjust the intervention according to users' behavior and responses, or both (FR34 to FR37). When this is the case, observers need access to follow-up components which give access to results (Figure 3G). Another type is that in which specialists design an intervention program to provide information or to send instructions to participants, as is the case with tutorial apps or self-care apps [59]. When this is the case, the specialists do not need information on how and when users interacted with the program, which demands the follow-up procedure to be optional.

ESPIM: Ontology-Based Conceptual Model

Our study contributes to a conceptual model (Figure 4) guiding specialists in adopting existing software platforms or building new ones along with a software development team. We represent the conceptual model using an ontology given its wide adoption [60-64] and description power [65].

Figure 4. ESPIM: Ontology-based Conceptual Model.



In the ESPIM ontology, the *Person* class represents observers (specialists) and participants (target users). The person concept may retrieve data from existing information systems using a unique key (eg, email). Participants may have relationships with 1 or more users in the system (eg, child, caregiver, partner).

The *Program* class represents an intervention program that encompasses a set of events that defines sets of active tasks or sensor data sampling (sensor-based sampling). Observers manage programs individually or collaboratively.

The *Event* class represents personalized or standardized intervention events applied by observers. An event comprises sets of tasks (active set) or sensor-based sampling. The *Active Set* class represents tasks required explicitly for a participant, represented by classes contained in the *Active Task* class. The *Sensor* class represents tasks achieved via sensor-based sampling, which may demand continuous and unobtrusive data collection at defined time intervals. An active set can be associated with sensor-based sampling occurring while the participant performs the tasks. Events initiate according to defined trigger conditions (in *Trigger* class).

The *Trigger* class encompasses trigger conditions and an overall set up that determines when and how an event is triggered. A trigger can hold the following features: self-initiated, temporal, contextual, random/reason, and specialist-initiated trigger. A self-initiated trigger allows target users to start an event, a temporal trigger schedules events based on time and dates, a contextual-based trigger considers context information obtained by sensors or by device usage, and random/reason-based triggers randomly deliver a defined number of intervention events during

XSL•FO
RenderX

a period. Observers may control triggers remotely. A trigger setup indicates notification timeout and obtrusiveness level.

The *Active Task* class corresponds to the delivery of stimulus that requires an interactive response. An active task contains at least one media stimulus (eg, text, image, audio, video). In the model, an active task comprises 4 types of stimuli: read message, question, media request, and external app.

In the *Active Task* class, the *Read Message* class represents sending multimedia messages to users. The *Question Active Task* class represents questions in different formats, including open-ended, multiple- and single-choice, and scales (eg, Likert). A question may include textual or other media elements. Choice questions represent loops and branches. A *Media Request Active Task* class represents the request for media assets (eg, audio). Finally, the *External App Active Task* class represents the activation of external apps, sending customized activation values and receiving completion results.

The *Media* class represents media stimuli employed in an active task instance. The model admits adding multiple media elements to a task.

The *Sensor* class represents the set up associated with sensor-based sampling: what should be collected (eg, social interaction, facial expressions) by which device (ie, wearable devices). Collection may occur over a period (sensor-based sampling) or during user interaction (active task).

The *Result* class represents data collected by the intervention program via active tasks or sensor-based sampling. Moreover, every participant interaction, or lack of interaction when expected, should be logged. Observers access instances of the *Results Session* class.

The *Condition* class represents actions triggered by conditional rules encompassing active (user interaction based) or passive (sensor based) data. For instance, if a participant fails to answer a notification, it is possible to execute an action such as alerting a particular observer or scheduling a new trigger.

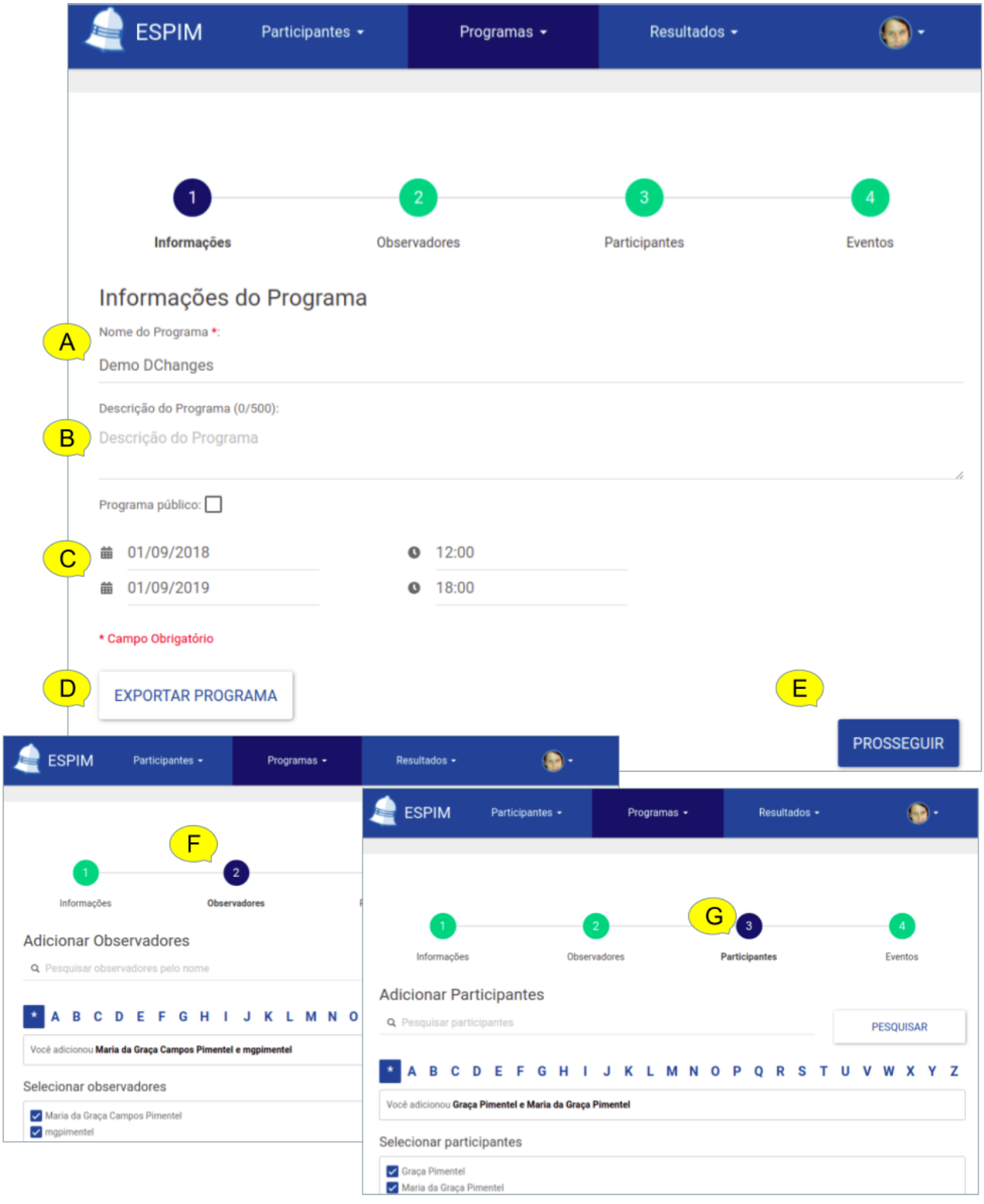
The *Phase* class represents the organization of an intervention program in stages. The *Phase* class aggregates events, and each event may be included in 1 or more phases. A condition that defines when a participant should proceed to another phase defines a phase duration. Conditions may be time, contextual, or response dependent. Moreover, the *Status* class in the ESPIM ontology allows registering a participant's progress in an intervention program that has phases.

ESPIM Software

The iterative design leading to the ESPIM method (Figure 1) involved the iterative prototyping of the software instance (employed by specialists when authoring; Figures 5-8) and monitoring (Figures 9 and 10) an intervention program, and the mobile app used by participants (Figure 11; see Multimedia Appendix 3 for details).

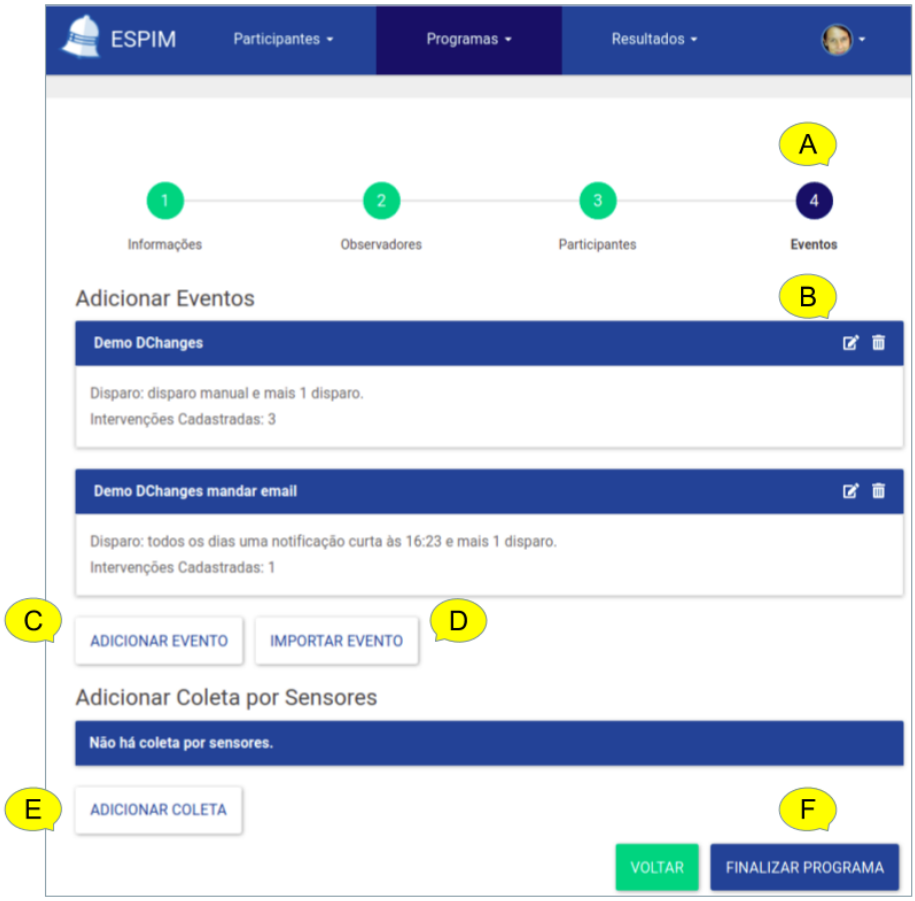
XSL•FO
RenderX

Figure 5. ESPIM web application used by the specialist to create an Intervention Program. The first step (A) informs the program's name and description (B), and duration (C). Options include exporting the program (D). The following steps register specialists (observers-F) and target-users (participants-G).



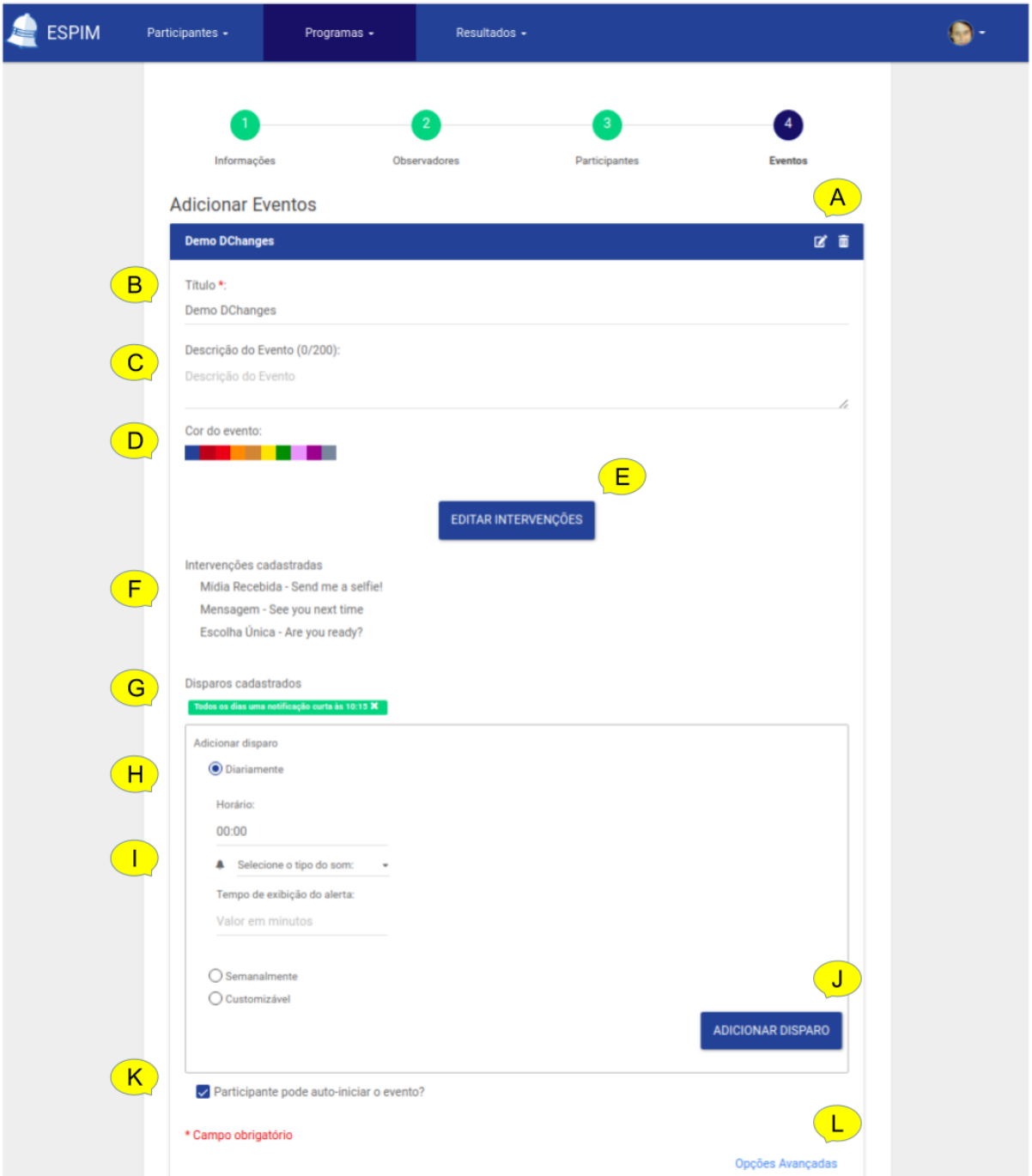
XSL•FO
RenderX

Figure 6. ESPIM web application step with options for Events. Options in this step (A) include editing an existing event (B), editing a new event from scratch (C), or by importing an existing one (D), specifying collection based on sensors (E). Finalizing is always available (F).



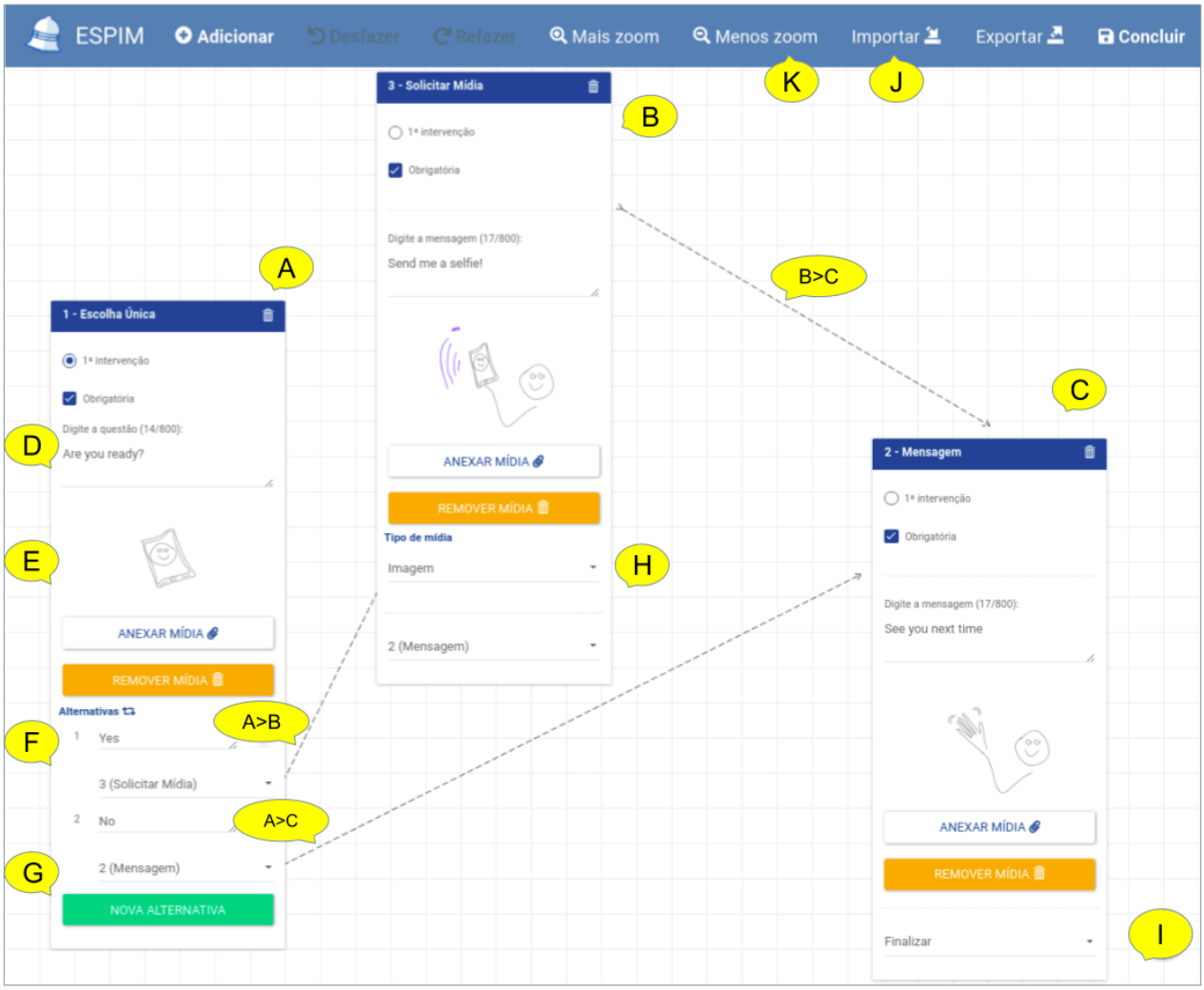
XSL•FO
RenderX

Figure 7. ESPIM web application step for editing one Event. For each Event (A), specialists provide a name (B), description (C), and color-coding (D). A button gives access to the interface for editing the corresponding flow of active tasks (E). This step shows the text from existing interventions (F) along with current triggers (G). Specialists create time-based triggers (H), and configure (I) and save the corresponding alarm types. Specialists can set self-initiated events (K) and configure that observers receive alerts when participants interact or miss an alarm (L).



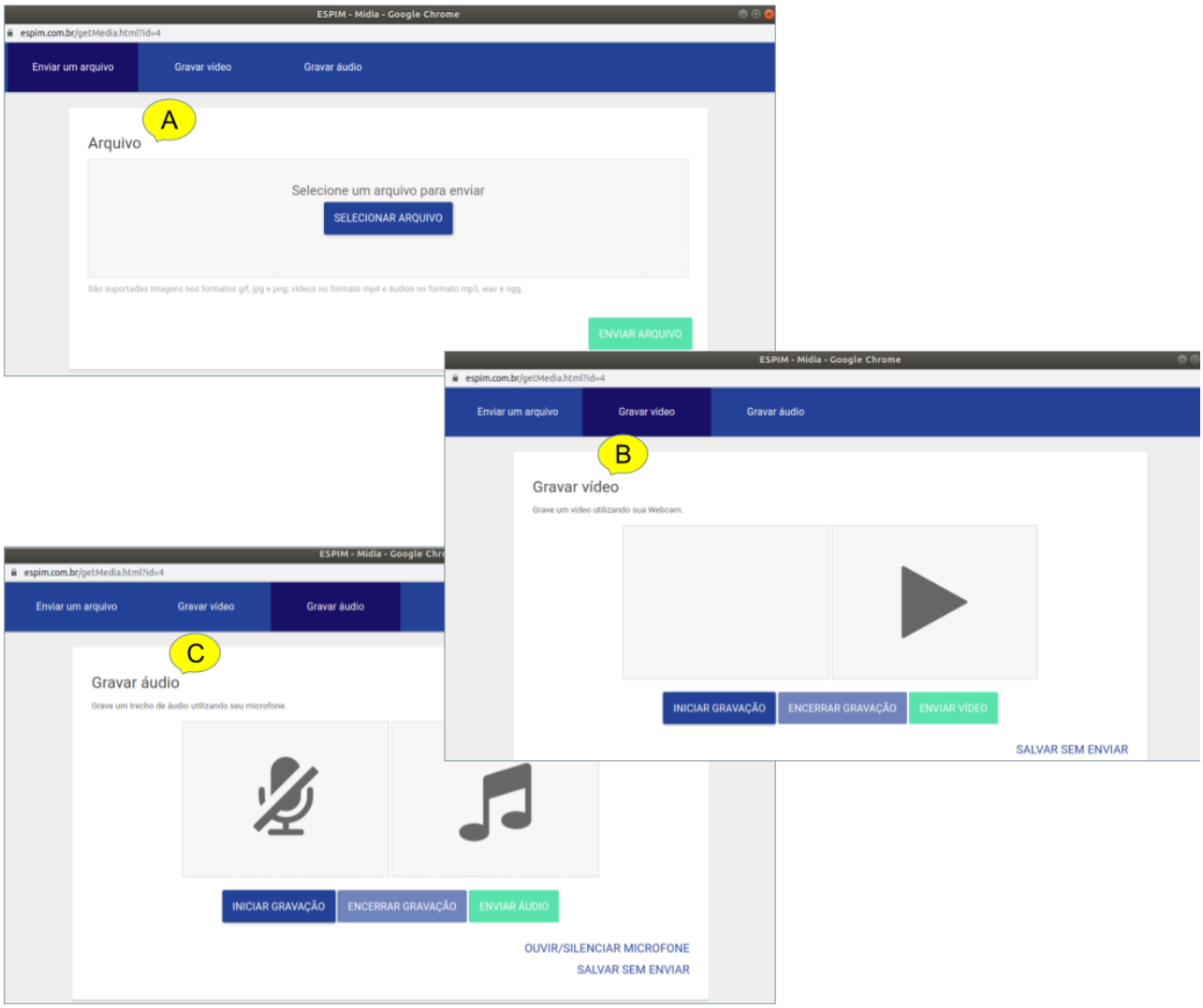
XSL•FO
RenderX

Figure 8. ESPIM web interface to create an Active Task Flow. This flow contains three interventions: a single-choice question (A), a task-based intervention requesting an image (B), and a message intervention (C). Specialists may include instructions using text (D) or other media (image (E), audio or video), or both. They must indicate the initial intervention (radio button in A), and mark each intervention as mandatory or optional (checkbox in A-C). The app shows arrows to indicate the flow (eg B>C). In a single-choice intervention (F), specialists may associate specific interventions to each alternative (A>B and A>C). Also, they can choose among many alternatives to choice- and scale-based questions (G). When specialists create a task requesting media (B), they indicate the type of media required ("image" in H). They must nominate at least one closing intervention (I). The specialist can zoom (K), and import and export (J) flows.



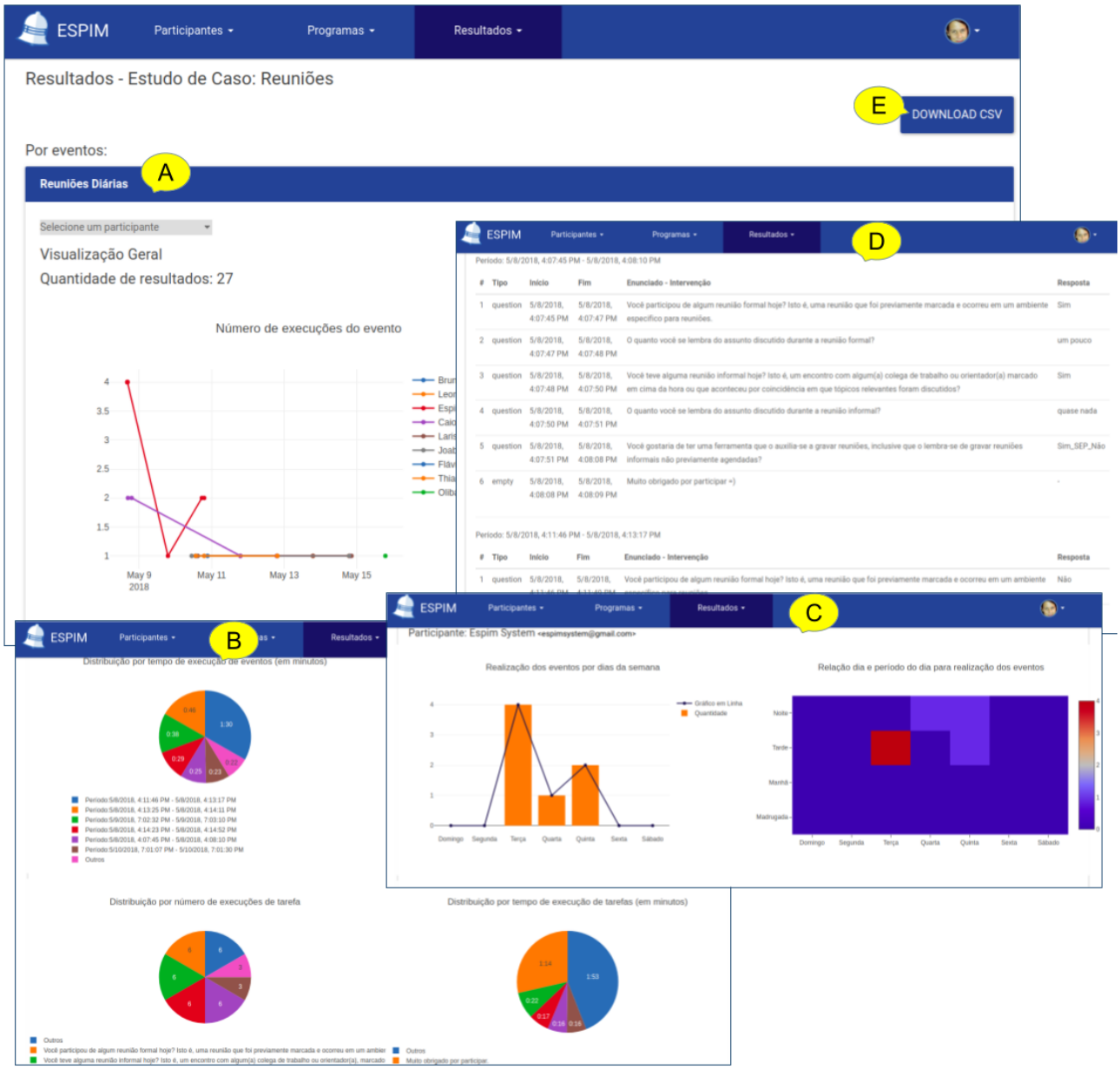
XSL•FO
RenderX

Figure 9. ESPIM web interface to upload media-based stimuli (A) and to record video (B) or audio (C).



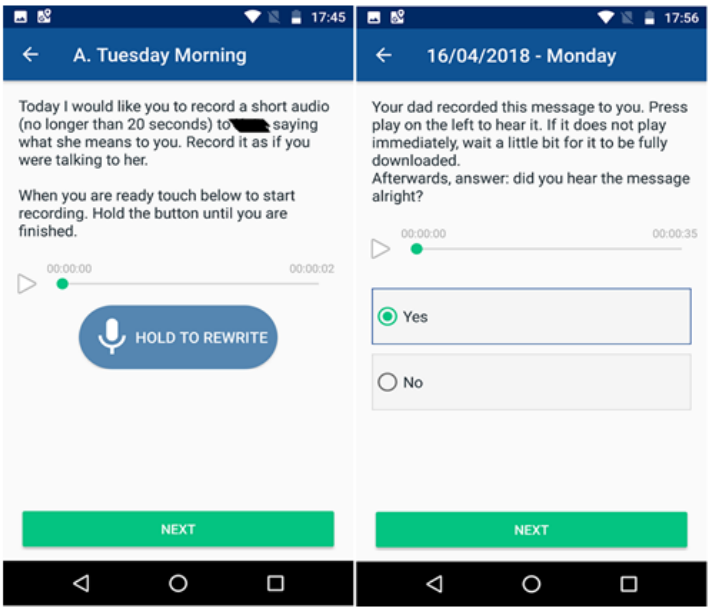
XSL•FO
RenderX

Figure 10. ESPIM web interface to visualize results includes an overview by a participant (A), distributions of responses both per task (B) and along the time (C), access to individual responses (D), and an option for download (E).



XSL•FO
RenderX

Figure 11. Mobile app presenting interventions. Active task requesting the user to record an audio message (left). Active task inviting the user to respond to a single-choice question upon listening to an audio stimulus (right).



Evaluation Outcomes

Interview (Predesign)

The predesign interviews lead to eliciting requirements from a group of specialists from several backgrounds (Figure 1C). We consolidate the results into ESPIM's functional requirements (Table 2).

First Heuristic Evaluation (App Prototype)

For the app prototype (Figure 1D), the HCI evaluation team reported the following issues: (1) navigation problems (store navigation and branching paths, confirm before quitting, clear paths upon abandonment/finishing); (2) inconsistent display on different versions of the operating system; (3) improve screen use (landscape orientation and fix overlay in small screens when the virtual keyboard is visible); (4) generic/uninformative notifications; and (5) nonstandard fonts. We fixed these issues and applied the material design guidelines [66].

First Usability Test (Web Application Prototype)

Each specialist interacted individually with the web application (Figure 1F) according to an 8-task hypothetical scenario (see Table SM1 in Multimedia Appendix 1). ESPIM researchers (including KRHR and IZ) observed the interaction. Upon conclusion, specialists answered the evaluation questionnaires and participated in a semistructured interview to report problems, feelings, and expectations. Results from each task are as follows.

Specialist P1 could not complete tasks 7 and 8 as she could not find the "Save" button. Specialists P3 and P5 partially completed task 7 (see Figure SM1 in Multimedia Appendix 1) as they added extra activities to the intervention planned. Regarding errors, specialists P2, P3, and P5 made 4 errors; P4 and P1 made 6 and 9 errors, respectively. The errors included creating an existing observer or participant, opening an external page during the interaction, using the first intervention element (a star icon) as a "Save" button, trying to save an empty program, and unexpected interactions with the interventions' flow screen.

Regarding the tasks, specialists made more than 1 error in tasks 4 and 7. In task 4, all specialists were uncertain about the meaning and function of the "observers" role. Task 7 was the one in which the specialists presented more difficulty and made more mistakes. This task had several substeps and demanded more attention, so specialists spent the longest (mean 13 min [SD 0.55]).

Results from the subjective evaluation of the interface elements range from –3 to +3. Specialists rated positively most aspects evaluated. Aspects evaluated +2 (Good) and +3 (Excellent) were available features, ease of memorization, and learning during use. The items with the lowest ratings were feedback/error messages, ease to undo errors, and information organization of the system. One negative score (–1) was attributed to "ease to undo errors" by P1 and was consistent with her overall performance. P1 could not fulfill 1 task and partially fulfilled another (tasks 8 and 7, respectively). P1 also made the highest number of errors.

The results from the UX while interacting with the system were positive, ranging between 1.4 and 3 (–3 to +3 range), meaning that specialists associated their experience with positive qualifiers, concepts, or feelings. The most positive evaluations were "efficiency" (mean 2.55 [SD 0.60]) and "dependability" (mean 2.35 [SD 0.87]). Evaluation of the remaining factors were as follows: attractiveness, mean 2.16 (SD 0.79); perspicuity, mean 2.10 (SD 0.78); stimulation, mean 2.10 (SD 0.71); and novelty, mean 2.05 (SD 0.68). As a result, specialists considered that they could interact efficiently, execute tasks reliably, understand and find functionalities, and felt stimulated and motivated during the interaction. They also perceived the system as innovative and attractive regarding interface elements.

XSL•FO
RenderX

Checklist

The UI/UX designers who designed the graphical interfaces of the ESPIM web applications suggested 9 improvements to the corresponding implementation (Figure 1J), including components behavior according to the material design guidelines, user lists presentation order, and text size and spacing. We adjusted them accordingly.

Heuristic Evaluation

For the heuristic evaluation of the ESPIM web application (Figure 1J2), the evaluators' report contained 71 usability issues. They classified 24 problems with severity 1 (cosmetic), 16 with severity 2 (minor), 26 with severity 3 (major), and 5 with severity 4 (catastrophic). Evaluators associated problems with 4 heuristics: consistency and standards, user control and freedom, flexibility and efficiency of use, and aesthetic and minimalist design. The ESPIM team promptly solved issues classified with severity 3 and 4 to provide a version for tests with gerontologists (Figure 1L1 and Figure 1L2). We fixed the other problems in later versions [52,67].

For the heuristic evaluation of the ESPIM mobile app (Figure 1J3), the evaluators' report comprised 21 usability issues. They classified 4 problems with severity 1 (cosmetic), 6 with severity 2 (minor), 8 with severity 3 (major), and 2 with severity 4 (catastrophic). Although evaluators associated 1 issue with the user control and freedom heuristic (7/21), we argued that we purposely designed some guiding-based features to make the app accessible to a wide range of user profiles. Yet, evaluators considered that several icons and some nomenclature lack intuitiveness and associated the issues with the correspondence between the system and the real-world heuristic (5/21). Evaluators associated other issues with the following heuristics: system status visibility, recognition instead of memorization, flexibility and efficiency of use, and aesthetic and minimalist design. We solved the problems before tests with gerontologists (Figure 1L1 and Figure 1L2).

Second Usability Test (Web Application) and Interview (App)

For the second usability test of the web application (Figure 1L1), each specialist interacted with the system according to a set of 10 tasks contemplated by 1 hypothetical scenario (see Table SM2 in Multimedia Appendix 1). ESPIM researchers (including KRHR and BCRC) observed the interaction. According to the protocol, the specialists answered the evaluation questionnaires and participated in a semistructured interview (see Textbox SM5 in Multimedia Appendix 1).

Specialists P1, P4, P6, and P5 completed all tasks. Specialists P2 and P3 partially completed task 7 (see Figure SM1 in Multimedia Appendix 1) because they did not find the requested media for uploading in the computer used for the test. Regarding errors, most happened at the planning interventions screen (Figure 7). Some specialists (3/6) faced difficulties in defining the intervention flow, especially the beginning and end.

Regarding the tasks, specialists made more than 1 error in tasks 4, 7, and 8. In task 4 (creating and adding a new observer), the specialists had difficulties registering emails in the interface. Task 7 was again the one in which the specialists made more mistakes. They spent more time executing task 7 (planning intervention flow; mean 15 min [SD 0.7]) as this task required more attention and the most steps. In task 8 (adding triggers), specialists faced difficulties understanding interface elements. For all other activities, the average execution time of participants was less than 2 min (SD 0.43).

Results from the questionnaire reporting subjective evaluation of interface elements were positive in most aspects. Aspects evaluated between "good" and "excellent" were layout of interface elements, available features, and ease of memorization. Five of six evaluators rated the aspect "learning during the use" between "excellent" and "very good." Specialist P2, who did not complete the tasks, evaluated this aspect as "too bad," reported "I really liked using ESPIM, it seems to me a very pleasant, intuitive tool" and offered the following suggestion, which we incorporated later: "[...] sometimes I lost a description explaining what the specific term meant: having short and clear information would have facilitated my journey."

At the concluding interviews, the specialists provided input as instructors or tutors in digital literacy courses for older people (CS-ElderlyDL; Figure 1H2). Specialists remarked as positive the simple aesthetic of the interface with "well-chosen colors" (colors considered color-blind users) and the in-app back button, justifying that some older people face difficulties finding the smartphone back button. Interviewees related that the app improved the engagement and learnability of the older people in the classes (especially when compared with control groups), multimedia messages promoted positive emotions, and mobile-delivered homework helped avoid evasion. Interviewees also reported some issues: users did not understand the asterisk in mandatory fields (consider disabling the next button), users did not know how to react to open questions (consider the automatic display of keyboard), users did not notice that a media was already captured (consider feedback with media preview), users clicked outside of multiple-choice options (consider exhibiting them inside inline boxes), lack of configuration for notification intrusiveness, and lack of in-app configuration for enabling/disabling 3G network usage.

Interview (Postdesign)

We conducted semistructured interviews with 8 specialists to identify the limitations they faced while conducting their case studies (Figure 1R). We classified the limitations as related to the model or the software. Model-related limitations imply modifying or extending the model to allow representing the required solution. Implementation-related limitations are restricted by the current system version: the model represents the corresponding solutions, but these are not yet available in the system. Because the latter are implementation specific, we discuss only model-related limitations.

Interviewees reported 4 limitations that require revisions in the conceptual model. The first limitation is related to the lack of support to represent the management of participants' records. Such a support would allow specialists to register information they consider relevant. To allow representing such a feature, we must extend the model accordingly.

XSL•FO
RenderX

The second limitation concerns representing the organization of participants into groups other than direct relationships among participants (as per FR10). The solution requires creating a class that allows multiple associations between specialists, participants, and groups so that a participant may be part of various groups managed by different specialists.

The other 2 limitations concern groups of events: we should extend the model to represent both hierarchical and customizable groups of events. The analogous implementation would allow the corresponding visual customization of groups of events.

Although these limitations require changes in the conceptual model, they did not limit the planning, authoring, or deploying of the interventions by the specialists.

Discussion

Principal Findings

Our study tackles limitations faced by health professionals determined to conduct mHealth interventions. Our main results are the ESPIM method and the conceptual model, which succeeded in leading specialists in planning, authoring, and deploying mHealth intervention programs with the support of a representative software system. ESPIM comprises (1) a list of requirements for mHealth experience sampling and intervention-based methods and systems, (2) a 4-dimension planning framework, (3) a 7-step-based process, and (4) an ontology-based conceptual model. A subsidiary result is the ESPIM software.

The list of requirements, offered to specialists when planning and deploying interventions, results from literature review, predesign interviews, iterative prototyping, and postdesign interviews. The 4-dimension planning framework aims to guide the planning of mHealth interventions by specialists by clarifying elements to be considered when using mobile, wearable, and ubiquitous technologies. The framework helps designing an intervention program while supporting team communication. The 7-step-based process guides the procedures that support ESPIM-based software.

A main contribution of ESPIM is a conceptual model aimed at guiding specialists in adopting existing software platforms or building new ones. The ESPIM ontology–based conceptual model represents the requirements collected and refined during our long-term design experience with health specialists and their research/professional needs. As a conceptual model, it defines a process and a common vocabulary to plan mobile device–mediated interventions; it also aids the development of mHealth apps.

Overall, ESPIM components can guide the planning and deployment of an intervention program using existing solutions, drive selecting one among the available tools, or guide the implementation of a novel specific or general platform.

While inspection-based usability evaluations of the ESPIM interfaces identified problems, we used the HCI specialists' feedback to make continuous improvements to the ESPIM system. Such evaluations preceded usability tests with professionals and their target users, and the resulting adjustments lead to intuitive and satisfactory usability tests.

Limitations

Limitations of our study are associated with those of the contributions. With respect to the list of requirements, even though requirements were elicited via interviews, iterative design, and empirical evaluation, some requirements might be left out, in particular due to the lack of case studies that demand supporting relationships among participants (FR10) or that employ home sensors, wearable devices, and JITAIs along with machine learning–based triggers [68,69]. These limitations are reflected in the 4-dimension planning framework, in the sensor dimension which might be more detailed to support such scenarios that demand, for instance, human activity recognition [70]. In future studies, new interviews can be conducted until new requirements are not identified [71]. In addition, we are running accessibility studies to provide solutions that can be used by different target-user profiles.

Concerning the 7-step-based process, the last step "results" should be extended to further assist data collection such as retrospective surveys [72], data analysis, and study reports. Besides, as data collection grows, integrating data analysis algorithms may support building predictive user models [69,73,74]. Further, with the growing adoption of experience sampling and program-based mHealth interventions, support to the specialists should be provided toward the production of reports in sufficient details to allow both replication and theory building, as was the case with randomized controlled trials [75].

Limitations of the ontology-based conceptual model include those identified in the postdesign interviews, which can be tackled by supporting participants' records and by the organization of participants into groups and the organizing events into hierarchical groups. These and other features can be promoted with the formalization of the current integration with external services [57].

Furthermore, while our work included usability evaluations to improve both the method and the system, dealing with barriers highlighted by Byambasuren et al [47], and provided a complementary method for JITAIs guidelines, as proposed by Nahum-Shani et al [76], evaluations of mHealth interventions still are a prominent gap in the literature which ESPIM did not address, as precisely identified by Bradway et al [77] and Dick et al [78].

Finally, our study identified important nonfunctional requirements which have not been included in the method's list of requirements, even though they are in consonance with recent literature and are partially attended by the current system. As a result of a systematic review on mHealth-related apps, Llorens-Vernet and Miró [79] offer a list of criteria, grouped into categories, aimed at guiding the development of mHealth apps. The systematic review ensured the categories (usability, privacy, security, appropriateness and suitability, transparency and content, safety, technical support and updates, and technology) are consistent with those identified by other authors in related contexts [7-10].

XSL•FO
RenderX

Comparison With Prior Work

The ESM and EMA methods for mobile-based data collection are a popular alternative [25,80], while EMIs and JITAIs depend on mobile technologies to monitor contextual changings. Some contributions report authoring and deployment systems.

MyExperience [81] pioneered employing ESM via a customizable mobile app, supporting sensor data capture, contextual triggers configuration, and questionnaires authoring. The authors predicted several scenarios and inspired later works. However, ESM programs were added to mobile devices as XML files, and real-time monitoring was not possible. MyExperience's spin-off movisensXS [32] is a platform with a graphical interface for creating questions with flow logic and media capture, temporal and contextual triggers, and real-time monitoring.

PACO (Personal Analytics COmpanion) [82] is an open-source system that allows the authoring of questions and media capture with branching features associated with logical operators. Despite its simplicity, PACO was the first to present an end-user graphical interface.

The ExperienceSampler [83] and AWARE [84] frameworks simplify the creation of ESM apps through logical and declarative programming. ExperienceSampler allows users to implement apps with messages, questions, skip and branching logic, and random-based notifications. AWARE focuses on logging sensor-based context information; its flexible contextual model considers 8 question types, branching logic, time, context-based notifications, and conditional broadcasts (eg, announce when each user answers a question).

Among commercially available systems, LifeData [31] enables the authoring of 14 types of tasks, including capturing photos and exhibiting a website, with logic and temporal flows, and self-initiated and random-based triggers. LifeData also enables the configuration of conditional actions and relationships among target users. Mobile EMA [33] provides a complex authoring interface, and allows displaying images in questions and integrating smartwatches to collect sensor data. LifeData and Mobile EMA systems comply with many of the functional requirements discussed in this paper, being extensively used in academic research [26-30].

Rough and Quigley [36] present recommendations and requirements for ESM authoring systems. They conducted interviews with 13 researchers and clinicians, 2 one-hour clinical observations, and 3 case studies with psychology researchers. They also propose the block-based visual programming tool Jeeves. The authors elicit 5 functional requirements for ESM tools: (1) collaboration and (2) support/share of projects, (3) tailoring of protocols and reminders to individuals, (4) debriefing/feedback/reminders in addition to surveys (ie, alternative types of tasks), and (5) ability to test both appearance and contextual behavior.

Investigating intervention design by specialists, Nahum-Shani et al [35] provided a framework for organizing theoretical and practical evidence into a model that supports authoring JITAIs. The authors discussed approaches for defining elements including states of vulnerability/opportunity and receptivity, outcomes, and adaptation strategy. Later, Nahum-Shani et al [76] proposed key components and design principles for designing mHealth JITAIs. The authors remark that authoring demands specialists to be concerned with components such as content, media, types of signal, temporal and contextual opportunities, receptivity marks, among others.

These works are complementary to ESPIM as they guide aspects of interventions' planning. ESPIM provides a step-by-step model and ontology that allows organizing mHealth intervention components considering requirements from ESM, EMA, EMIs, JITAIs, and long-term studies with specialists from diverse areas. Furthermore, the ESPIM system demonstrates the feasibility of developing an authoring system based on the method, encompassing imperative features and providing flexibility for integrating ecological and just-in-time interventions.

Conclusions

The ESPIM comprises a list of requirements for mHealth experience sampling and intervention-based methods and systems, a 4-dimension planning framework, a 7-step-based process, and an ontology-based conceptual model. The ESPIM system encompasses web and mobile apps. Besides overseeing the planning of an intervention program, ESPIMs components guide the design of an ESPIM-based software platform as in our study. Moreover, current limitations point to further research as well as practical actions.

Eight case studies show that the ESPIM method and system allowed specialists to be the users who planned, created, and deployed interventions. The case studies encompassed interventions by professionals from psychology, gerontology, computer science, speech therapy, and occupational therapy. Specialists' target users were parents of children diagnosed with autism spectrum disorder, older persons, university students, children, and older person's caregivers. The specialists reported being able to create and conduct their studies without modifying their original design. A qualitative evaluation of the ontology-based conceptual model showed its compliance to the elicited functional requirements.

Acknowledgments

The São Paulo Research Foundation (FAPESP) supported IZ (Grant Numbers 2017/09549-6 and 2015/18117-7), EANS (Grant Number 2018/14674-7), and MGCP (Grant Numbers 2016/00351-6 and 2016/50489-4). The Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) supported BCRC (Grant Number 150465/2016-5) and MGCP (Grant Number 312058/2015-2). The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) supported the study via Finance Code 001. We thank all participants in the case studies, interviews, and evaluations.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluations protocols and tasks.
[DOC File , 1374 KB - jmir_v23i7e24278_app1.doc]

Multimedia Appendix 2
Semi-structured interview questions.
[DOC File , 32 KB - jmir_v23i7e24278_app2.doc]

Multimedia Appendix 3
ESPIM system details.
[DOC File , 1028 KB - jmir_v23i7e24278_app3.doc]

References

1. Byambasuren O, Beller E, Glasziou P. Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
2. Alwashmi MF, Fitzpatrick B, Davis E, Gamble J, Farrell J, Hawboldt J. Perceptions of Health Care Providers Regarding a Mobile Health Intervention to Manage Chronic Obstructive Pulmonary Disease: Qualitative Study. JMIR Mhealth Uhealth 2019 Jun 10;7(6):e13950 [FREE Full text] [doi: 10.2196/13950] [Medline: 31199330]
3. Weichelt B, Bendixsen C, Patrick T. A Model for Assessing Necessary Conditions for Rural Health Care's Mobile Health Readiness: Qualitative Assessment of Clinician-Perceived Barriers. JMIR Mhealth Uhealth 2019 Nov 08;7(11):e11915 [FREE Full text] [doi: 10.2196/11915] [Medline: 31702564]
4. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
5. Lingg M, Lütschg V. Health System Stakeholders' Perspective on the Role of Mobile Health and Its Adoption in the Swiss Health System: Qualitative Study. JMIR Mhealth Uhealth 2020 May 11;8(5):e17315 [FREE Full text] [doi: 10.2196/17315] [Medline: 32391802]
6. Lupiáñez-Villanueva F, Folkvord F, Vanden Abeele M. Influence of the Business Revenue, Recommendation, and Provider Models on Mobile Health App Adoption: Three-Country Experimental Vignette Study. JMIR Mhealth Uhealth 2020 Jun 04;8(6):e17272 [FREE Full text] [doi: 10.2196/17272] [Medline: 32496204]
7. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' Role in the Adoption of an Oncology Decision Support App in Europe and Its Implications for Organizational Practices: Qualitative Case Study. JMIR Mhealth Uhealth 2019 May 03;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
8. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding Clinicians' Adoption of Mobile Health Tools: A Qualitative Review of the Most Used Frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
9. Jacob C, Sanchez-Vazquez A, Ivory C. Social, Organizational, and Technological Factors Impacting Clinicians' Adoption of Mobile Health Tools: Systematic Literature Review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
10. Ramdani B, Duan B, Berrou I. Exploring the Determinants of Mobile Health Adoption by Hospitals in China: Empirical Study. JMIR Med Inform 2020 Jul 14;8(7):e14795 [FREE Full text] [doi: 10.2196/14795] [Medline: 32459630]
11. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. J Youth Adolesc 1977 Sep;6(3):281-294. [doi: 10.1007/BF02138940] [Medline: 24408457]
12. Larson R, Csikszentmihalyi M, Graef R. Mood variability and the psychosocial adjustment of adolescents. J Youth Adolescence 1980 Dec;9(6):469-490. [doi: 10.1007/bf02089885]
13. Csikszentmihalyi M, Larson R. Validity and reliability of the Experience-Sampling Method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [doi: 10.1097/00005053-198709000-00004] [Medline: 3655778]
14. Stone A, Shiffman S. Ecological Momentary Assessment (Ema) in behavioral medicine. Annals of Behavioral Medicine 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
15. Wang L, Miller L. Just-in-the-Moment Adaptive Interventions (JITAI): A Meta-Analytical Review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
16. Pereira AAC, Destro JR, Picinin Bernuci M, Garcia LF, Rodrigues Lucena TF. Effects of a WhatsApp-Delivered Education Intervention to Enhance Breast Cancer Knowledge in Women: Mixed-Methods Study. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17430 [FREE Full text] [doi: 10.2196/17430] [Medline: 32706726]

XSL•FO
RenderX

17. Bogale B, Mørkrid K, O'Donnell B, Ghanem B, Abu Ward I, Abu Khader K, et al. Development of a targeted client communication intervention to women using an electronic maternal and child health registry: a qualitative study. BMC Med Inform Decis Mak 2020 Jan 06;20(1):1 [FREE Full text] [doi: 10.1186/s12911-019-1002-x] [Medline: 31906929]
18. Champoux E, Price R, Cowdery JE, Dinh M, Meurer WJ, Rehman N, et al. Reach Out Emergency Department: Partnering With an Economically Disadvantaged Community in the Development of a Text-Messaging Intervention to Address High Blood Pressure. Health Promot Pract 2020 Sep;21(5):791-801. [doi: 10.1177/1524839920913550] [Medline: 32228238]
19. Odukoya O, Faseru B, Uguru N, Jamda M, Onigbogi O, James O, et al. Text messaging promoting physician-led brief intervention tobacco cessation: A before-and-after study among physicians in three tertiary hospitals in Nigeria. Subst Abus 2020;41(2):186-190 [FREE Full text] [doi: 10.1080/08897077.2018.1526844] [Medline: 30373478]
20. Abroms L, Whittaker R, Free C, Mendel Van Alstyne J, Schindler-Ruwisch J. Developing and Pretesting a Text Messaging Program for Health Behavior Change: Recommended Steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
21. van Dijk MR, Oostingh EC, Koster MPH, Willemsen SP, Laven JSE, Steegers-Theunissen RPM. The use of the mHealth program Smarter Pregnancy in preconception care: rationale, study design and data collection of a randomized controlled trial. BMC Pregnancy Childbirth 2017 Jan 26;17(1):46 [FREE Full text] [doi: 10.1186/s12884-017-1228-5] [Medline: 28125970]
22. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong KL, et al. Assessing User Engagement of an mHealth Intervention: Development and Implementation of the Growing Healthy App Engagement Index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
23. Quercia K, Tran PL, Jinoro J, Herniainasolo JL, Viviano M, Vassilakos P, et al. A Mobile Health Data Collection System for Remote Areas to Monitor Women Participating in a Cervical Cancer Screening Campaign. Telemed J E Health 2018 Apr;24(4):277-282. [doi: 10.1089/tmj.2017.0146] [Medline: 28846504]
24. Bramwell SE, Meyerowitz-Katz G, Ferguson C, Jayaballa R, McLean M, Maberly G. The effect of an mHealth intervention for titration of insulin for type 2 diabetes: A pilot study. Eur J Cardiovasc Nurs 2020 Jun;19(5):386-392. [doi: 10.1177/1474515119889764] [Medline: 31736378]
25. van Berkel N, Ferreira D, Kostakos V. The Experience Sampling Method on Mobile Devices. ACM Comput. Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]
26. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018 Jul 10;35(7):601-608. [doi: 10.1002/da.22730] [Medline: 29637663]
27. Shrier LA, Burke PJ, Kells M, Scherer EA, Sarda V, Jonestrask C, et al. Pilot randomized trial of MOMENT, a motivational counseling-plus-ecological momentary intervention to reduce marijuana use in youth. Mhealth 2018 Jul;4:29-29 [FREE Full text] [doi: 10.21037/mhealth.2018.07.04] [Medline: 30148142]
28. O'Connor SG, Habre R, Bastain T, Toledo-Corral C, Gilliland F, Eckel S, et al. Within-subject effects of environmental and social stressors on pre- and post-partum obesity-related biobehavioral responses in low-income Hispanic women: protocol of an intensive longitudinal study. BMC Public Health 2019 Feb 28;19(1):253 [FREE Full text] [doi: 10.1186/s12889-019-6583-x] [Medline: 30819155]
29. Bell IH, Rossell SL, Farhall J, Hayward M, Lim MH, Fielding-Smith SF, et al. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): Feasibility, acceptability and preliminary clinical outcomes. Schizophr Res 2020 Feb;216:479-487. [doi: 10.1016/j.schres.2019.10.026] [Medline: 31812327]
30. Yang YS, Ryu GW, Park CG, Yeom I, Shim KW, Choi M. Mood and Stress Evaluation of Adult Patients With Moyamoya Disease in Korea: Ecological Momentary Assessment Method Using a Mobile Phone App. JMIR Mhealth Uhealth 2020 May 25;8(5):e17034 [FREE Full text] [doi: 10.2196/17034] [Medline: 32449687]
31. LifeData Experience Sampling App for Research. URL: https://www.lifedatacorp.com/ [accessed 2021-06-17]
32. Experience Sampling - movisensXS - movisens GmbH. URL: https://www.movisens.com/en/products/movisensxs/ [accessed 2021-06-18]
33. Mobile EMA. URL: https://ilumivu.com/solutions/ecological-momentary-assessment-app/ [accessed 2021-06-18]
34. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
35. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
36. Rough DJ, Quigley A. End-User Development of Experience Sampling Smartphone Apps -Recommendations and Requirements. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2020 Jun 15;4(2):1-19. [doi: 10.1145/3397307]
37. Schuler D, Namioka A. Participatory Design: PrinciplesPractices. 1st ed. New jersey: L. Erlbaum Associates Inc; Mar 01, 1993:978.

XSL•FO
RenderX

38. Spinuzzi C. The methodology of participatory design. Technical Communication Internet Society for Technical Communication 2005 May 1;52(2):163-174 [FREE Full text]
39. Nielsen J, Molich R. Heuristic evaluation of user interfaces. New York: Association for Computing Machinery; 1990 Mar 01 Presented at: Conference on Human Factors in Computing Systems; 04/01/1990; Seattle p. 249-256.
40. Nielsen J. Usability inspection methods. Conference on Human Factors in Computing Systems - Proceedings Internet New York, New York, USA: Association for Computing Machinery; 1994 Apr 24 Presented at: Conference on Human Factors in Computing Systems; 04/24/1994; Boston.
41. Carroll JB, Osgood CE, Suci GJ, Tannenbaum PH. The Measurement of Meaning. Language 1959 Jan;35(1):58. [doi: 10.2307/411335]
42. Laugwitz B, Held T, Schrepp M. Construction and Evaluation of a User Experience Questionnaire. Springer, Berlin, Heidelberg 2008:63-76. [doi: 10.1007/978-3-540-89350-9_6]
43. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
44. Camilli G, Vargas S, Ryan S, Barnett WS. Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. Teachers College Record 2010;112(3):579-620 [FREE Full text]
45. Lozano-Lozano M, Fernández-Lao C, Cantarero-Villanueva I, Noguerol I, Álvarez-Salvago F, Cruz-Fernández M, et al. A Blended Learning System to Improve Motivation, Mood State, and Satisfaction in Undergraduate Students: Randomized Controlled Trial. J Med Internet Res 2020 May 22;22(5):1-13. [doi: 10.2196/17101]
46. Watts TW, Bailey DH, Li C. Aiming Further: Addressing the Need for High-Quality Longitudinal Research in Education. Journal of Research on Educational Effectiveness 2019 Dec 06;12(4):648-658. [doi: 10.1080/19345747.2019.1644692]
47. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and Facilitators of the Prescription of mHealth Apps in Australian General Practice: Qualitative Study. JMIR Mhealth Uhealth 2020 Jul 30;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
48. Zaine I, Rodrigues KRH, Cunha BCR, Viel CC, Orlando AF, Machado OJN, et al. ESPIM: An Ubiquitous Data Collection and Programmed Intervention System using ESM and Mobile Devices. 2016 Nov 08 Presented at: Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web - Webmedia '16 Internet .–14; 2016; Teresina. [doi: 10.1145/2976796.2988222]
49. Rodrigues KRG, Cunha BCR, Zaine I, Viel CC, Scalco LF, Pimentel MGC. ESPIM System: Interface Evolution to Enable Authoring and Interaction with Multimedia Intervention Programs. ESPIM System: Interface Evolution to Enable Authoring and Interaction with Multimedia Intervention Programs. Anais Principais do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia) Internet SBC; 2018 Oct 16 Presented at: Brazilian Symposium on Multimedia and the Web; 10/16/2018; Salvador URL: https://sol.sbc.org.br/index.php/webmedia/article/view/4565 [doi: 10.1145/3243082.3243120]
50. Cunha B, Rodrigues K, Zaine I, Scalco L, Viel C, Maria DGC. Web-based authoring of multimedia intervention programs for mobile devices: a case study on elderly digital literacy. USA: Association for Computing Machinery; 2019 Apr 08 Presented at: The 34th ACM/SIGAPP Symposium on Applied Computing; 04/08/2019; Limassol. [doi: 10.1145/3297280.3297325]
51. Cunha BCR, Rodrigues KRH, Pimentel MGC. Synthesizing guidelines for facilitating elderly-smartphone interaction. New York: Association for Computing Machinery, Inc; 2019 Oct 29 Presented at: Brazilian Symposium on Multimedia and the Web; 11/29/2019; Rio de Janeiro. [doi: 10.1145/3323503.3349563]
52. Rodrigues KRH, Viel CC, Zaine I, Cunha BCR, Scalco LF, Pimentel MGC. Data Collection and Intervention Personalized as Interactive Multimedia Documents. 2017 Oct 17 Presented at: Brazillian Symposium on Multimedia and the Web; 10/17/2017; Gramado. [doi: 10.1145/3126858.3131574]
53. Cunha BCR, Rodrigues KRH, Scalco LF, Zaine I, Viel CC, Pimentel MGC. Authoring of interventions with on-demand generation of screen for mobile devices. 2018 Oct 16 Presented at: Brazilian Symposium on Multimedia and the Web; 10/16/2018; Salvador. [doi: 10.5753/webmedia.2018.4588]
54. Zaine I, Benitez P, Rodrigues KDH, Pimentel MDG. Applied Behavior Analysis in Residential Settings: Use of a Mobile Application to Support Parental Engagement in At-Home Educational Activities. Creative Education 2019 Aug 19:1883-1903. [doi: 10.4236/ce.2019.108136]
55. Cachioni M, Zaine I, Chiarelli TM, Cliquet LOBV, Rodrigues KRH, Cunha BCR, et al. Aprendizagem ao longo de toda a vida e letramento digital de idosos: um modelo multidisciplinar de intervenção com o apoio de um aplicativo. RBCEH 2019 Aug 20;16(1):18-24. [doi: 10.5335/rbceh.v16i1.9751]
56. Zaine I, Frohlich DM, Rodrigues KRH, Cunha BCR, Orlando AF, Scalco LF, et al. Promoting Social Connection and Deepening Relations Among Older Adults: Design and Qualitative Evaluation of Media Parcels. J Med Internet Res 2019 Oct 03;21(10):1-18 [FREE Full text] [doi: 10.2196/14112] [Medline: 31584001]
57. Viel CC, Rodrigues KRH, Cunha BCR, Pimentel MGC. Elderly vloggers: authoring and sharing stories with mobile media and DTV using the the GhostWriter service. 2019 Oct 29 Presented at: Brazilian Symposium on Multimedia and the Web; 10/29/2019; Rio de Janeiro URL: https://sol.sbc.org.br/index.php/webmedia/article/view/8020 [doi: 10.5753/webmedia_estendido.2020]

XSL•FO
RenderX

58. Harari G, Lane N, Wang R, Crosier B, Campbell A, Gosling S. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. Perspect Psychol Sci 2016 Nov 28;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
59. Liu K, Xie Z, Or C. Effectiveness of Mobile App-Assisted Self-Care Interventions for Improving Patient Outcomes in Type 2 Diabetes and/or Hypertension: Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2020 Aug 04;8(8):e15779 [FREE Full text] [doi: 10.2196/15779] [Medline: 32459654]
60. Kramarz B, Lovering R. Gene Ontology: A Resource for Analysis and Interpretation of Alzheimer's Disease Data. Alzheimer's Disease Internet 2019:23-36. [doi: 10.15586/alzheimersdisease.2019.ch2] [Medline: 31895510]
61. Shi Z, Traverso A, van Soest J, Dekker A, Wee L. Technical Note: Ontology-guided radiomics analysis workflow (O-RAW). Med Phys 2019 Dec 25;46(12):5677-5684 [FREE Full text] [doi: 10.1002/mp.13844] [Medline: 31580484]
62. Parreiras F, Walter T, Gröner G. Visualizing ontologies with UML-like notation. New York, NY: Association for Computing Machinery; 2010 Oct Presented at: ODiSE'10: Ontology-Driven Software Engineering; October 2010; Reno, NV p. 1-6. [doi: 10.1145/1937128.1937132]
63. Dermeval D, Vilela J, Bittencourt I, Castro J, Isotani S, Brito P, et al. Applications of ontologies in requirements engineering: a systematic review of the literature. Requirements Eng 2015 Feb 14;21(4):405-437. [doi: 10.1007/s00766-015-0222-6]
64. Harrow I, Balakrishnan R, Jimenez-Ruiz E, Jupp S, Lomax J, Reed J, et al. Ontology mapping for semantically enabled applications. Drug Discovery Today 2019 Oct 01;24(10):2068-2075. [doi: 10.1016/j.drudis.2019.05.020]
65. Verdonck M, Gailly F, de Cesare S, Poels G. Ontology-driven conceptual modeling: A systematic literature mapping and review. Applied Ontology 2015 Dec 04;10(3-4):197-227. [doi: 10.3233/ao-150154]
66. Material Design. URL: https://material.io/ [accessed 2021-06-24]
67. Viel CC, Rodrigues KRH, Zaine I, Cunha BCR, Scalco LF, Pimentel MGC. Personalized Ubiquitous Data Collection and Intervention as Interactive Multimedia Documents. 2017 Sep 04 Presented at: ACM Symposium on Document Engineering 2017; 09/04/2017; Valletta. [doi: 10.1145/3103010.3121046]
68. Ghosh S, Ganguly N, Mitra B, De P. Designing An Experience Sampling Method for Smartphone based Emotion Detection. IEEE Trans. Affective Comput 2019 Mar 18:1-1. [doi: 10.1109/taffc.2019.2905561]
69. Kapoor A, Horvitz E. Experience sampling for building predictive user models: A comparative study. New York: ACM Press; 2008 Apr 05 Presented at: CHI Conference on Human Factors in Computing Systems; 04/05/2008; Florence. [doi: 10.1145/1357054.1357159]
70. Zimmermann LC, Rodrigues KRH, Pimentel MGC. EPARS: elderly physical activity reminder system using smartphone and wearable sensors. New York: Association for Computing Machinery; 2019 Sep 09 Presented at: The 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 09/09/2019; London. [doi: 10.1145/3341162.3350845]
71. Gerhart N. Generalizing stopping rule research: Development of scales. Journal of Computer Information Systems 2018 Feb 08;60(1):93-100. [doi: 10.1080/08874417.2017.1416315]
72. Naab TK, Karnowski V, Schlütz D. Reporting Mobile Social Media Use: How Survey and Experience Sampling Measures Differ. Communication Methods and Measures 2018 Dec 25;13(2):126-147. [doi: 10.1080/19312458.2018.1555799]
73. Pryss R, Schlee W, Hoppenstedt B, Reichert M, Spiliopoulou M, Langguth B, et al. Applying Machine Learning to Daily-Life Data From the TrackYourTinnitus Mobile Health Crowdsensing Platform to Predict the Mobile Operating System Used With High Accuracy: Longitudinal Observational Study. J Med Internet Res 2020 Jun 30;22(6):e15547 [FREE Full text] [doi: 10.2196/15547] [Medline: 32602842]
74. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
75. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):1-10 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
76. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
77. Bradway M, Gabarron E, Johansen M, Zanaboni P, Jardim P, Joakimsen R, et al. Methods and Measures Used to Evaluate Patient-Operated Mobile Health Interventions: Scoping Literature Review. JMIR Mhealth Uhealth 2020 Apr 30;8(4):1-17 [FREE Full text] [doi: 10.2196/16814] [Medline: 32352394]
78. Dick S, O'Connor Y, Thompson MJ, O'Donoghue J, Hardy V, Wu TJ, et al. Considerations for Improved Mobile Health Evaluation: Retrospective Qualitative Investigation. JMIR Mhealth Uhealth 2020 Jan 22;8(1):1-10 [FREE Full text] [doi: 10.2196/12424] [Medline: 32012085]
79. Llorens-Vernet P, Miró J. Standards for Mobile Health-Related Apps: Systematic Review and Development of a Guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):1-9 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]

XSL•FO
RenderX

80. Stamate D, Katrinecz A, Alghamdi W, Stahl D, Delespaul P, van OJ, et al. Predicting psychosis using the experience sampling method with mobile apps. 2017 Dec 18 Presented at: 16th IEEE International Conference on Machine Learning and Applications; 12/18/2017; Cancun. [doi: 10.1109/icmla.2017.00-84]
81. Froehlich J, Chen MY, Consolvo S, Harrison B, Landay JA. MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. New York: ACM Press; 2007 Jun 11 Presented at: The Fifth International Conference on Mobile Systems, Applications, and Services; 06/11/2007; San Juan. [doi: 10.1145/1247660.1247670]
82. Baxter KK, Avrekh A, Evans B. Using experience sampling methodology to collect deep data. New York: Association for Computing Machinery; 2015 Apr 18 Presented at: CHI Conference on Human Factors in Computing Systems; 04/18/2015; Seoul. [doi: 10.1145/2702613.2706668]
83. Thai S, Page-Gould E. ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. Psychol Methods 2018 Dec 01;23(4):729-739. [doi: 10.1037/met0000151] [Medline: 28616998]
84. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile Context Instrumentation Framework. Front. ICT 2015 Apr 20;2:1-9. [doi: 10.3389/fict.2015.00006]

Abbreviations

EMA: ecological momentary assessment
EMI: ecological momentary intervention
ESM: experience sampling method
ESPIM: Experience Sampling and Programmed Intervention Method
FR: functional requirements
HCI: human–computer interaction
JITAI: just-in-time adaptive intervention
mHealth: mobile health
PACO: Personal Analytics COmpanion
UEQ: User Experience Questionnaire
UI: user interface
UX: user experience

Edited by R Kukafka; submitted 11.09.20; peer-reviewed by C Jacob, R Krukowski, Z Ren; comments to author 21.09.20; revised version received 09.10.20; accepted 25.02.21; published 12.07.21.

Please cite as:
Cunha BCR, Rodrigues KRDH, Zaine I, da Silva EAN, Viel CC, Pimentel MDGC
Experience Sampling and Programmed Intervention Method and System for Planning, Authoring, and Deploying Mobile Health Interventions: Design and Case Reports
J Med Internet Res 2021;23(7):e24278
URL: https://www.jmir.org/2021/7/e24278
doi:10.2196/24278
PMID:34255652

©Bruna Carolina Rodrigues Cunha, Kamila Rios Da Hora Rodrigues, Isabela Zaine, Elias Adriano Nogueira da Silva, Caio César Viel, Maria Da Graça Campos Pimentel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Text Mining Techniques to Identify Health Care Providers With Patient Safety Problems: Exploratory Study

Iris Hendrickx[1], MA, PhD; Tim Voets[1], MA; Pieter van Dyk[2], MSc, PhD; Rudolf B Kool[3], MD, PhD

[1]Centre for Language Studies, Centre for Language and Speech Technology, Faculty of Arts, Radboud University, Nijmegen, Netherlands

[2]Dutch Health and Youth Care Inspectorate, Utrecht, Netherlands

[3]IQ healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Iris Hendrickx, MA, PhD
Centre for Language Studies, Centre for Language and Speech Technology
Faculty of Arts
Radboud University
Erasmusplein 1
Nijmegen, 6525HT
Netherlands
Phone: 31 2436 15575
Email: i.hendrickx@let.ru.nl

Abstract

Background: Regulatory bodies such as health care inspectorates can identify potential patient safety problems in health care providers by analyzing patient complaints. However, it is challenging to analyze the large number of complaints. Text mining techniques may help identify signals of problems with patient safety at health care providers.

Objective: The aim of this study was to explore whether employing text mining techniques on patient complaint databases can help identify potential problems with patient safety at health care providers and automatically predict the severity of patient complaints.

Methods: We performed an exploratory study on the complaints database of the Dutch Health and Youth Care Inspectorate with more than 22,000 written complaints. Severe complaints are defined as those cases where the inspectorate contact point experts deemed it worthy of a triage by the inspectorate, or complaints that led to direct action by the inspectorate. We investigated a range of supervised machine learning techniques to assign a severity label to complaints that can be used to prioritize which incoming complaints need the most attention. We studied several features based on the complaints' written content, including sentiment analysis, to decide which were helpful for severity prediction. Finally, we showcased how we could combine these severity predictions and automatic keyword analysis on the complaints database and listed health care providers and their organization-specific complaints to determine the average severity of complaints per organization.

Results: A straightforward text classification approach using a bag-of-words feature representation worked best for the severity prediction of complaints. We obtained an accuracy of 87%-93% (2658-2990 of 3319 complaints) on the held-out test set and an F1 score of 45%-51% on the severe complaints. The skewed class distribution led to only reasonable recall (47%-54%) and precision (44%-49%) scores. The use of sentiment analysis for severity prediction was not helpful. By combining the predicted severity outcomes with an automatic keyword analysis, we identified several health care providers that could have patient safety problems.

Conclusions: Text mining techniques for analyzing complaints by civilians can support inspectorates. They can automatically predict the severity of the complaints, or they can be used for keyword analysis. This can help the inspectorate detect potential patient safety problems, or support prioritizing follow-up supervision activities by sorting complaints based on the severity per organization or per sector.

(J Med Internet Res 2021;23(7):e19064) doi:10.2196/19064

KEYWORDS

text mining; risk management; health care quality improvement

XSL•FO
RenderX

Introduction

We know since some time that information from patients can help improve the quality and safety of care [1]. Previous studies [2-4] using patient and client experiences with health care providers focused on reviews in which patients describe their positive and negative experiences. By analyzing patient complaints, health care providers can detect preventable patient safety issues with opportunities for improvement [5], enable organizational learning, and identify poor outcomes [6,7]. We also know that a small group of health care providers causes a major part of these complaints [8]. Identifying these providers with potential safety problems is important to significantly improve patient safety [9].

Regulatory bodies such as health care inspectorates already use negative patient experiences to identify health care providers with patient safety problems. In Australia, health complaints commissions predict the risks of individual doctors becoming the subject of repeated patient complaints [10]. In the Netherlands, the Health and Youth Care Inspectorate uses repeated negative ratings from patient rating websites as part of their supervision to predict risks [11,12]. Furthermore, the Dutch Health and Youth Care Inspectorate started a national contact point for health care complaints in 2014. All citizens can file complaints against a health care provider at this national contact point of the inspectorate over telephone, using an online web form, or via email. The complaints are categorized and manually judged based on severity by the inspectorate contact point. In the majority of the complaints, the contact point can offer straightforward advice to the patients by referring them a local complaints officer, patient counselor, or the relevant external dispute resolution body of the health care provider. Incoming complaints from the public are only passed on to the national health care inspectors when the complaints are deemed to be extremely serious such as major patient safety risks, fraud, misconduct, and sexual harassment. Approximately 13% of the complaints are triaged by inspectors [13]. All complaints are stored in a database, which is a relatively new source of information for the inspectorate. However, the database comprehends huge amounts of data and it is challenging to identify the potential safety problems in this heap of mostly unstructured information. Text mining can be defined as "analyzing patterns in text data to extract and discover actionable knowledge directly useful for task completion or decision-making, thus providing more direct task support for users" [14].

As incoming complaints are handled one by one by the contact point team members, certain providers may have received several complaints, which on their own are not sufficiently severe to warrant further inspection. However, a series of complaints may indicate safety problems. Given the successful use of machine learning techniques in extracting information from large quantities of data, the application of these techniques may be promising for identifying health care providers with patient safety problems in such complaint databases.

The aim of this study was to explore whether the application of text mining techniques on the complaints database of the Dutch Health and Youth Care Inspectorate could help identify health care providers with patient safety problems that may harm patients. These problems are especially within the scope of the Dutch inspectorate. We investigated whether we could support the inspectorate by using text mining techniques to automatically predict the severity of the complaints and conducting sentiment analysis to prioritize visiting health care providers with potential patient safety problems.

Methods

Data Collection and Preparation

Since 2014, the Dutch health care system has a national contact point for complaints about care designated by the Dutch Health and Youth Care Inspectorate. All the incoming complaints are registered in a specific database with structured information formats. Each entry contains several fields such as the personal details of the person filing the complaint and the targeted health care provider. In this study, we focused on the request fields, providing descriptions of the complaints filed by the citizens. Other fields are also manually added by the contact point members such as category labels, sector labels, priority labels, reporting dates, whether the complaints are sent to an inspector for triage, and action fields listing the actions taken by the inspectorate. In total, the database contains 22,509 complaints for the time period between 2014 and 2017. Complaints arrive at the inspectorate contact point via different media. Telephone calls are transcribed by a contact point team member as a written summary of the verbal complaint, whereas web forms are linked directly to the database fields. Email content is usually added to the request field, which may also include notes from possible follow-up contacts. Letters, constituting the minority media, are scanned to a digital format and a brief summary is added by the contact point team in the "short description" field. In 2017, two-thirds of the complaints were filed over telephone and a quarter using online web forms.

As the complaint data are highly sensitive, during our research, the data remained on a secure server of the inspectorate that was not connected to the Internet and could only be accessed via secure login. Therefore, we had to bring the text mining software to the data. We designed a ready-made research environment for the text mining experiments [15] in the form of a virtual machine that was then installed on the local secure server.

We extracted the complaints from the inspectorate's database and attempted to remove all mentions of personal information such as telephone numbers, addresses, and names from the free-text fields containing the complaint descriptions using regular expressions. We used the fields with personal information to detect the names and numbers to remove from the free-text fields in a preprocessing step. Removing every name and address reference in a text automatically is practically infeasible [16], but most personal information was removed. The motivation for removing person names from the text was for privacy reasons and obtaining clean data for text mining. The removed unique names and numbers are uninformative attributes in the text mining process. As the data were always kept on the secure inspectorate server, privacy was already

XSL•FO
RenderX

guaranteed. The study was assessed and approved by the Research Ethics Committee of the Radboud University Medical Center in the Netherlands. The ethics committee waived the request to approve the study, as it did not fall under the Medical Research Involving Human Subjects Act in the Netherlands (number 2020-7024).

Severity Prediction

Some complaints are more severe than others and the concept of severity can be viewed as a sliding scale. Gillespie and Reader [17] have also been labeling severity in health care complaints. Their Healthcare Complaints Analysis Tool (HCAT) is a well-validated instrument for manual labeling of complaint severity. It is a taxonomy and coding scheme for manual annotation of health care complaints. Each complaint is analyzed based on seven problem categories (such as quality communication and safety), and the level of severity (low, medium, and high) is assessed. In this HCAT coding scheme, the problem severity is coded separately from the health care outcomes (harm). The HCAT tool is intended for complaint labeling by human experts. However, we practically conceptualized severity as a supervised machine learning task, namely learning predict the severity of unseen complaints from the manual judgments on the severity of complaints. As mentioned in the introduction, the majority of the complaints are handled by the contact point members by offering advice or by redirecting them to local complaint committees or counselors. Only potential major patient safety risks are redirected to the national health care inspectors. We investigated two variants of the manually assigned labels in the Dutch complaints database that can be considered indicators of severity. First, every complaint that was sent from the contact point to an inspector for triage (Triage) was considered severe. Second, using a more restrictive option, we only considered complaints to be severe when the inspectorate decided to take action (such as investigating the health care provider) based on the complaint (Decision). Approximately 13% percent of all the complaints were sent to the inspectorate for triage, and approximately 6% were further investigated after the triage. We experimented with these two indicators to determine whether we could automatically learn to predict severity based on the written content of the complaints with a supervised machine learning approach.

Information Representation

An obvious use of automatic severity prediction is that such an automatic technique can be applied to the incoming digital complaints to rank them based on the predicted severity so that the contact point members can prioritize the most severe complaints. For incoming complaints, only the request field (description of the complaints filed by the citizens) is known, and we used that field as the main information source of the complaint description. We represented the information from the request field in five ways:

Bag-of-Words Representation

The simplest representation of the text in the request field is to create a bag-of-words representation that contains a list of n-grams (a sequence of n neighboring words). The value of n was varied from 1 (single word) to a maximum length of 3. We restricted the list of n-grams to only those n-grams that occur in at least 5 complaints. We experimented with two different weighting schemes (termed frequency–inverse document frequency weighting and log scaling) to determine the importance of the n-grams. We also varied the list size of the top selected most informative words.

Specified Keywords

Instead of a bag of words, one can also focus on a subset of keywords that are expected to be informative. We filtered words in the complaints by mapping them to the Linguistic Inquiry and Word Count (LIWC) dictionary designed for psycholinguistic research [18]. This is accomplished by extracting keywords from the text and grouping them into categories. Each of these categories corresponds to a particular concept. This can include grammatical concepts (whether a word is a pronoun, verb, noun, etc) as well as concepts such as emotional states, motivations, intentions, and thought processes [19]. The version used here was a Dutch version developed by Zijlstra et al [20] that has been independently verified. Mapping the keywords resulted in a total of 64 features per text. This Dutch version provides the full specifications of the features.

Word Embeddings

Word embeddings are a technique to represent word semantics on a high level (distributional semantics). It is based on the observation that words occurring in the same context [21] have similar meanings and this is captured at an abstract level by word embeddings. We applied Word2Vec (Google) [22] to implement word embeddings and mapped the vocabulary to a semantic space with 300 dimensions.

Document Attributes

Besides the content words in the documents, other document characteristics can also be automatically measured, namely sentence complexity, relational coherence in the text, and writing style indicators such as usage of action verbs and pronouns. Such textual characteristics could indicate emotions and writing styles may contain useful information that may indicate severity regardless of the actual content words in the text. We explored these feature types and used a list of 250 different document characteristics such as probability features including word, n-gram, and lemma frequencies; complexity features including sentence, word, and noun phrase lengths; and a wide range of other features. We computed these features with T-scan [23], a tool that was designed to predict the readability of a document.

Average Sentiment

We also used automatic sentiment analysis for feature representation, as we expected that severe complaints could contain more negative emotions.

Sentiment analysis is generally conducted in two ways; the first is to use a subjectivity lexicon that has annotated entries for various words, whereas the second involves classifying the documents for positive or negative sentiments using machine learning techniques. We applied the first technique and used a subjectivity lexicon for Dutch adjectives [24] to estimate the

XSL•FO
RenderX

overall sentiment value per complaint based on the text description in the request field.

Machine Learning Techniques

Given the exploratory nature of our work, we could not determine a priori which machine learning algorithm would perform the best on the data set. As is well known in the field of machine learning, and eloquently worded in the "No Free Lunch" theorem [25], there is not one machine learning algorithm that always provides the best solution. Therefore, we conducted experimental tests on the following algorithms to investigate which one was suitable for this particular task: multinomial naive Bayes, support vector machine (SVM), k-nearest neighbor (k-nn), and extreme gradient decision tree. We used the majority class baseline as the reference.

Parameters, Evaluation Metrics, and Experimental Setup

To estimate the performance of the classifier, we trained a model on one part of the data set and evaluated the performance on a held-out sample. We used the complaints filed in the first 6 months of 2017 as the testing material and used all other complaints as the training material.

We conducted experiments to optimize the parameter settings for each of the machine learning algorithms. We performed a 10-fold cross-validation/grid search on the training set to find the optimal combination of features, algorithms, and algorithmic parameter settings. The best classifier was then applied to the test set.

Note that the complaint data set had a skewed class balance, as only 6% of the complaints were labeled "severe" when using the strict "Decision" (The inspectorate took action on the basis of the complaint) label. Machine learning techniques are known to be sensitive to such class imbalances [26].

As the evaluation metrics to estimate the performance of the classifier, we computed recall, precision, and their harmonized mean, the F1 score [27], for the severity class label. This implies that we only focused on how well we performed on the minority class label "severe." We computed how many of the complaints that were actually labeled severe were also predicted by the system as "severe" (recall) and how many of the severity predictions were also actually labeled by humans as "severe" (precision).

As a second measure, we also computed the area under the curve-receiver operating characteristics (AUC-ROC) scores that show the balance between the true positive (TP) rate versus the false positive (FP) rate. This score also considers the predictions on the "not severe" class label.

Classifier Optimization

Each classifier was optimized by 10-fold cross-validation experiments on the training set. We experimented with each of the 5 different feature groups (bag of words, keywords, word embeddings, document attributes, and sentiment) individually, and each of the textual representations combined with the sentiment analysis features to investigate whether the sentiment features were predicted for severity. Lastly, we conducted an experiment combining all the feature representations. We ran grid search optimization experiments to find a suitable algorithmic parameter setting for each machine learning algorithm. We optimized the F1 score of the positive class because we were interested in a classifier that could predict the severity label.

These tuning experiments on the training set determined which classifier and which features worked best for predicting the severity of the complaints. The optimization experiments showed that the k-nn algorithm did not perform very well in this task with F1 scores of 24.5 (Triage) and 10.1 (Decision) and that the naive Bayes obtained the best results with F1 scores of 46.9 (Triage) and 32.1 (Decision).

The sentiment features did not contribute to better performance and were not helpful for severity prediction. The textual representation with the best result was the bag-of-words representation; it scored better than any of the other individual features and was also better than the combined feature representation. The classifier that performed the best on the training set (naive Bayes with the bag-of-words representation) was applied to the held-out test set. More details on the optimization experiments can be found in the supplement.

Identifying Health Care Providers With Patient Safety Problems

We explored automatic methods to predict the severity of a complaint not only so that the Health and Youth Care Inspectorate can prioritize the most urgent complaints but also to open up possibilities for using automatic techniques that can look for patterns in the current complaints database to spot cases with elevated risk levels. As all complaints are handled using a one-by-one strategy, methods that explore the entire database can provide new insights into the data.

We also showcased how these severity predictions could be combined with keyword patterns to quickly provide insights for the inspectorate on how to extract safety problem indicators per health care provider based on the database with complaints from several years. We performed an exploratory analysis for every health care provider for which at least 10 complaints were registered in the database and used the severity predictions to rank these organizations based on their level of urgency for further inspection. The content of these grouped complaints was represented as n-grams with the most typical words and phrases so that the inspectors could identify the topics at a glance. These most descriptive n-grams per health care provider were identified using a statistical metric, log likelihood, which compares the scores of the specific terms related to health care providers with those in the entire complaints database.

Results

Severity Prediction

Table 1 shows the F1 values and harmonic means of precision and recall computed for the "severe" class on the held-out test set. This is a strict measurement, as we have a skewed class distribution, and the obtained F1 scores are in line with the expectations for such skewed classes. The ROC-AUC scores

XSL•FO
RenderX

of 0.7 (Triage) and 0.8 (Decision) clearly indicate that the classifier performs well above chance level (random predictions lead to an AUC score of 0.5). Accuracy includes the correct predictions of the "not severe" cases, and therefore, the classifier attains high accuracy in both the labeling tasks. We can observe a slightly better score on the Triage label than the Decision label. The class distribution between severe and not severe is skewed, as only 6% of the complaints in the test set were labeled severe based on the Decision labels and 13% based on the Triage labels.

Table 1. Best results obtained by the naive Bayes on the test set.

Label	Accuracy	F1	Precision	Recall	AUC-ROC[a] score
Triage	86.9	51.1	48.8	53.7	0.72
Decision	93.0	45.2	43.6	46.8	0.81

[a]AUC-ROC: area under the curve-receiver operating characteristics.

We present the confusion matrices in Table 2. Recall reflects the TP and false negative (FN), whereas precision focuses on the TP and FP. Most of the complaints are correctly labeled "not severe," namely true negative (TN), which leads to the highly accurate scores shown in Table 1 and is also reflected in the AUC-ROC scores that consider the TN, as it reflects the TP rate against the FP rate. The skewed class distribution leads to only reasonable F1, recall, and precision scores. This is evident when observing the TP, FP, and FN values in this table. Our model overpredicts the "severe" class by mislabeling the 'not severe' complaints (FP) and overpredicts the "not severe" label for severe complaints (FN), leading to only a reasonable overall performance.

Table 2. Confusion matrices from which the scores in were computed.

Cell value		Triage		Decision	
TN[a]	FP[b]	2658	238	2990	124
FN[c]	TP[d]	196	227	109	96

[a]TN: true negative.

[b]FP: false positive.

[c]FN: false negative.

[d]TP: true positive.

Identifying Healthcare Providers With Patient Safety Problems

Table 3 shows the top selection of (anonymous) organizations for which multiple complaints were registered in the database. We sorted these health care providers based on their average severity score, the last column in Table 3. This was computed by dividing the number of predicted severity labels (shown in the column "Severity prediction") by the number of complaints (the second column). Note that the third and fourth columns, "Triage" and "Decision," represent the number of complaints that were actually triaged or inspected by the inspectors. The first row of Table 3 shows that health care provider 1 received 11 complaints of which 3 were triaged, and in 2 of these 3 cases, the inspectorate took action to place the health care provider under inspection.

Table 3. Number of complaints and severity per organization.

Health care provider	Number of complaints	Triage	Decision	Severity prediction	Average triage	Average decision	Average severity
1	11	3	2	*9*	27.3	18.2	*81.8*
2	10	3	2	*7*	30.0	20.0	*70.0*
3	11	0	0	*7*	0.0	0.0	*63.6*
4	19	4	2	*12*	21.1	10.5	*63.2*
5	72	12	7	*45*	16.7	9.7	*62.5*
6	13	3	1	*8*	23.1	7.7	*61.5*
7	10	4	4	*6*	40.0	40.0	*60.0*
8	10	3	2	*6*	30.0	20.0	*60.0*

The anonymized version of the 10 most important (translated) terms that were identified for health care provider 1 indicate the contents of the complaints concerning perceived patient mistreatment. The top 10 word n-grams for health care provider 1 are the following: sister, scolding, safety sister, harrowing, care sister, quality life, diet, note, neglect, and employee.

XSL•FO
RenderX

Discussion

Our study has explored whether supervised machine learning techniques can automatically determine the severity of incoming complaints. Our results showed that severity was best predicted with a straightforward text classification approach using the bag-of-words feature representation. We combined the severity predictions with word n-grams to create an overview of the most urgent complaints per individual health care provider. An overview based on the severity of the complaints could help the inspectorate in prioritizing health care providers with potential patient safety problems.

The sentiment features were not helpful in predicting the severity of the complaints. A possible explanation for this is that we are still far from achieving accurate automatic sentiment predictions [28] and that the sentiment scores were not sufficiently reliable. On the other hand, our hypothesis that severe complaints contain more negative emotions could be at fault as well. This aspect needs further investigation.

Comparison With Other Studies

Greaves and colleagues [3] showed that text mining techniques can be used to label online patient reviews of health care providers as positive or negative, and to extract information about the opinions on aspects of care quality. Several studies [4,29,30] use text mining techniques like text classification, topic modeling, and sentiment detection on online health care provider reviews to explore which topics are indicative of perceived health care quality by patients. We focused on labeling the severity of the complaints, whereas these studies focused on patient opinions and sentiments using fine-grained doctor visit–related topics in the online reviews. An interesting path for future work would be to extend the current severity label to a more fine-grained severity label that combines severity with the type of complaint to create an automatic method similar to that established by Brereton et al [31] who performed a manual qualitative analysis of negative reviews to identify the most frequent actionable criticisms in patient reviews of hospices.

Lui et al [32] combined a data-driven topic modeling approach with expert knowledge to create a topic taxonomy of medical practitioner reviews. Their analysis of a large data set of reviews also showed that patients with different diseases focus on other topics. In our current study, we did not diversify our results for different sectors, but this would certainly be one of the aspects to investigate in future work.

Desmet [33] applied machine learning techniques to determine severity based on written text by labeling severity in social media posts related to suicidal thoughts. This study represented severity using three labels (low, intermediate, and high); SVM and k-nn algorithms were used for classification and a range of different features such as bag of words, emotion lexicons, and clustered topical words. The severity of suicidal thoughts in social media posts was predicted with an F1 score of 43% to 67%.

Our study used machine learning techniques to estimate the severity of the complaints filed with the inspectorate. The Dutch Health and Youth Care Inspectorate has already incorporated patient opinions in its supervision of health care providers. Since 2016, the ratings of Zorgkaart Nederland, the national patient rating site, are being shown in the information dashboard for inspectors, which was seen as useful extra information by inspectors [8]. In the United Kingdom, The Care Quality Commission has experimented with identifying risks and prioritizing visits by combining patient feedback from the National Health Service (NHS) Choices, Patient Opinion, Facebook, and Twitter to obtain a near real-time collective judgment score for acute hospitals and trusts on any given date [34]. This so-called Patient Voice Tracking System was successful in identifying a high-risk group of organizations for inspection.

Implications for Practice

Our analysis can help inspectorates evaluate their own severity categorization. Moreover, the Dutch Health and Youth Care Inspectorate can use our algorithms to prioritize the incoming complaints directly based on an automatically predicted severity label. Furthermore, they can create an overview of all the complaints about a health care provider using our algorithm with an automatically predicted severity label. These labels can be used to identify soft signals [35] and blind spots [17] that indicate potential safety issues with certain health care providers. These are cases that on their own do not warrant direct action but may together indicate systematic problems. Our study produced some encouraging results, and currently, the inspectorate is investigating opportunities for using such algorithms as an assistant tool for the human experts in evaluating complaints and prioritizing visits.

Using text mining techniques such as n-grams in the set of provider-specific complaints can support prioritizing follow-up supervision activities by sorting complaints based on severity per organization or per sector. Health care inspectorates in most countries deal with several providers, and they cannot visit all the providers. Prioritizing visits based on severity prediction could improve the effectiveness of their supervision.

Predicting the severity of complaints based on algorithms can also be useful for health care providers themselves. The board of directors of the providers should be interested in knowing whether the trends of patient complaints indicate patient safety problems requiring improvement and thus prevent the inspectorate from taking action.

Strengths and Weaknesses of This Study

One strength of this study was the collaboration with the inspectorate, as this provided a different perspective to study complaints for detecting potential safety risks, whereas most of the previous works discussed above have focused on text mining reviews to extract information on patient satisfaction. The inspectorate also provided the database containing 22,000 complaints spanning 3 years. Another strength was the use of software-supported natural language processing for Dutch. This software is part of the open-source software La Machine [36]. This software could be installed on the secure server of the inspectorate, which enabled thorough analysis. A major weakness of our study was the difficult preprocessing phase. It took months to create a secure machine learning environment

XSL•FO
RenderX

that was able to host our software. For some complaints, it was difficult to automatically remove all forms of personal information.

Implications for Further Research

In our study, we used a general LIWC list of words, which can be replaced by a specific word list created by the inspectorate. This may improve the results.

We also tried to optimize the F1 score, and the harmonic means of recall (how many relevant complaints are selected?) and precision (how many selected complaints are relevant?). For practical use, it may be more logical to optimize a severity classifier concerning completeness to select the most serious cases at the expense of precision.

For using our method in practice, a thoughtful implementation study is necessary with inspectors in the lead to optimize their support. Further, sound and up-to-date technical infrastructure for data science is indispensable for this approach.

Conclusion

A fully automatic analysis of a complaints database with text mining techniques may support health care inspectorates in identifying potential patient safety problems at health care providers. In particular, a straightforward text classification approach using the bag-of-words feature representation can be effective for severity prediction. Using text mining techniques such as n-grams in the set of provider-specific complaints can support prioritizing follow-up supervision activities by sorting complaints based on the severity per organization or per sector.

Acknowledgments

This study was funded by the Dutch organization ZonMw (project 516004614).

Conflicts of Interest

None declared.

References

1. Vincent C, Coulter A. Patient safety: what about the patient? BMJ Qual Saf 2002 Mar;11(1):76-80 [FREE Full text] [doi: 10.1136/qhc.11.1.76]
2. Maramba ID, Davey A, Elliott MN, Roberts M, Roland M, Brown F, et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. JMIR Med Inform 2015 May 06;3(2):e20 [FREE Full text] [doi: 10.2196/medinform.3783] [Medline: 25947632]
3. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
4. Rivas R, Montazeri N, Le NX, Hristidis V. Automatic classification of online doctor reviews: evaluation of text classifier algorithms. J Med Internet Res 2018 Nov;20(11):e11141 [FREE Full text] [doi: 10.2196/11141] [Medline: 30425030]
5. Bismark MM. Relationship between complaints and quality of care in New Zealand: a descriptive analysis of complainants and non-complainants following adverse events. Qual Saf Health Care 2006 Feb;15(1):17-22. [doi: 10.1136/qshc.2005.015743] [Medline: 16456205]
6. Gillespie A, Reader TW. The Healthcare Complaints Analysis Tool: development and reliability testing of a method for service monitoring and organisational learning. BMJ Qual Saf 2016 Jan;25(12):937-946. [doi: 10.1136/bmjqs-2015-004596] [Medline: 26740496]
7. Råberus A, Holmström I, Galvin K, Sundler A. The nature of patient complaints: a resource for healthcare improvements. Int J Qual Health Care Oct 2019 Aug;31(7):556-562. [doi: 10.1093/intqhc/mzy215] [Medline: 30346537]
8. Bismark M, Spittal M, Gurrin L, Ward M, Studdert D. Identification of doctors at risk of recurrent complaints: a national study of healthcare complaints in Australia. BMJ Qual Saf 2013 Jul;22:532-540 [FREE Full text] [doi: 10.1136/bmjqs-2012-001691] [Medline: 23576774]
9. Paterson R. Not so random: patient complaints and 'frequent flier' doctors. Int J Qual Health Care 2013 Apr;22(7):525-527. [doi: 10.1136/bmjqs-2013-001902] [Medline: 30346537]
10. Spittal MJ, Bismark MM, Studdert DM. The PRONE score: an algorithm for predicting doctors' risks of formal patient complaints using routinely collected administrative data. BMJ Qual Saf 2015 Apr 08;24(6):360-368. [doi: 10.1136/bmjqs-2014-003834]
11. Kleefstra SM, Zandbelt LC, Borghans I, de Haes HJ, Kool RB. Investigating the Potential Contribution of Patient Rating Sites to Hospital Supervision: Exploratory Results From an Interview Study in the Netherlands. J Med Internet Res 2016 Jul;18(7):e201 [FREE Full text] [doi: 10.2196/jmir.5552] [Medline: 27439392]
12. Kool RB, Kleefstra SM, Borghans I, Atsma F, van de Belt TH. Influence of intensified supervision by health care inspectorates on online patient ratings of hospitals: a multilevel study of more than 43,000 online ratings. J Med Internet Res 2016 Jul;18(7):e198 [FREE Full text] [doi: 10.2196/jmir.5884] [Medline: 27421302]
13. LMZ. Klachtbeeld 2018. Utrecht, Netherlands: Inspectie Gezondheidszorg en Jeugd; May 2019.

XSL•FO
RenderX

14. Zhai C, Massung S. Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining. San Rafael, California: Association for Computing Machinery and Morgan & Claypool; 2016.
15. van Gompel M, Hendrickx I. LaMachine: A meta-distribution for NLP software. In: Selected papers from the CLARIN Annual Conference 2018. Linköping: Linköping University Electronic Press; 2018 Presented at: CLARIN Annual Conference; October 8-10, 2018; Pisa p. 214-226.
16. Narayanan A, Shmatikov V. Myths and fallacies of "personally identifiable information". Commun. ACM 2010 Jun;53(6):24-26. [doi: 10.1145/1743546.1743558]
17. Gillespie A, Reader TW. Patient-centered insights: using health care complaints to reveal hot spots and blind spots in quality and safety. Milbank Q 2018 Sep;96(3):530-567. [doi: 10.1111/1468-0009.12338]
18. Pennebaker J, Booth R, Francis M. Operator's Manual Linguistic Inquiry and Word Count: LIWC2007. Austin, Texas: LIWC.net; 2007.
19. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec;29(1):24-54. [doi: 10.1177/0261927x09351676]
20. Zijlstra H, van Middendorp H, van Meerveld T, Geenen R. Validiteit van de Nederlandse versie van de linguistic inquiry and word count (LIWC). NETP 2018 Aug;60(3):50-58. [doi: 10.1007/bf03062342]
21. Firth J. Reprinted in: Palmer, F. R. Selected Papers of J. R. Firth. In: A Synopsis of Linguistic Theory 1930-1955. London: Longmans; 1968.
22. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. 2013 Presented at: NIPS'13:Proceedings of the 26th International Conference on Neural Information Processing Systems; December 05, 2013; Nevada p. 3111-3119.
23. Pander MH, Kraf R, van den Bosch A, Dekker N, van Gompel M, Kleijn S. T-Scan: a new tool for analyzing Dutch text. Computational Linguistics in the Netherlands Journal 2014;4:53-74.
24. De Smet T, Daelemans W. Vreselijk mooi!(terribly beautiful): A subjectivity lexicon for Dutch adjectives. Instanbul: European Language Resources Association (ELRA); 2012 Presented at: The Eighth International Conference on Language Resources and Evaluation (LREC'12); May 2012; Istanbul p. 3568-3572.
25. Wolpert D, Macready W. No free lunch theorems for optimization. IEEE Trans Evol Computat 1997;1(1):67-82. [doi: 10.1109/4235.585893]
26. Chawla NV, Japkowicz N, Kotcz A. Editorial:special issue on learning from imbalanced data sets. ACM SIGKDD Explorations Newsletter 2004 Jun;6(1):1-6. [doi: 10.1145/1007730.1007733]
27. van Rijsbergen CJ. Information Retrieval. London: Buttersworth; 1979.
28. Ribeiro FN, Araújo M, Gonçalves P, André Gonçalves M, Benevenuto F. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci 2016 Jul;5(1):1-29. [doi: 10.1140/epjds/s13688-016-0085-1]
29. James T, Calderon E, Cook D. Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. Expert Syst Appl 2017 Apr;71:479-492 [FREE Full text] [doi: 10.1016/j.eswa.2016.11.004]
30. Doing-Harris KM, Mowery DL, Daniels C, Chapman WW, Conway M. Understanding patient satisfaction with received healthcare services: a natural language processing approach. AMIA Annu Symp Proc 2016:524-533 [FREE Full text] [Medline: 28269848]
31. Brereton EJ, Matlock DD, Fitzgerald M, Venechuk G, Knoepke C, Allen LA, et al. Content analysis of negative online reviews of hospice agencies in the United States. JAMA Netw Open 2020 Feb;3(2):e1921130. [doi: 10.1001/jamanetworkopen.2019.21130]
32. Li J, Liu M, Li X, Liu X, Liu J. Developing embedded taxonomy and mining patients' interests from web-based physician reviews: mixed-methods approach. J Med Internet Res 2018 Aug;20(8):e254 [FREE Full text] [doi: 10.2196/jmir.8868] [Medline: 30115610]
33. Desmet B. Finding the online cry for help,automatic text classification for suicide prevention. In: PhD Thesis. Ghent: Ghent University; 2014.
34. Griffiths A, Leaver MP. Wisdom of patients: predicting the quality of care using aggregated patient feedback. BMJ Qual Saf 2018 Feb;27(2):110-118 [FREE Full text] [doi: 10.1136/bmjqs-2017-006847] [Medline: 28971881]
35. Wallenburg I, Kok J, Bal R. Omgaan met Soft Signals in het Toezicht: Signaleren, Interpreteren en Duiden van Risico's in de Zorg door de IGJ. Rotterdam: Erasmus University Rotterdam; Feb 2019.
36. van Gompel M. LaMachine software distribution. URL: https://proycon.github.io/LaMachine/ [accessed 2021-01-05]

Abbreviations

AUC-ROC: area under the curve-receiver operating characteristics
FN: false negative
FP: false positive
HCAT: Healthcare Complaints Analysis Tool
k-nn: k-nearest neighbor
LIWC: Linguistic Inquiry and Word Count

XSL•FO
RenderX

NHS: National Health Service
SVM: support vector machine
TN: true negative
TP: true positive

Edited by R Kukafka; submitted 02.04.20; peer-reviewed by R Boumans, S McLennan; comments to author 12.06.20; revised version received 31.08.20; accepted 13.05.21; published 27.07.21.

Please cite as:
Hendrickx I, Voets T, van Dyk P, Kool RB
Using Text Mining Techniques to Identify Health Care Providers With Patient Safety Problems: Exploratory Study
J Med Internet Res 2021;23(7):e19064
URL: https://www.jmir.org/2021/7/e19064
doi:10.2196/19064
PMID:34313604

©Iris Hendrickx, Tim Voets, Pieter van Dyk, Rudolf B Kool. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Challenges and Pitfalls of Detecting Sleep Hypopnea Using a Wearable Optical Sensor: Comparative Study

Zhongxing Zhang[1,2], PhD; Ming Qi[1], MSc; Gordana Hügli[1], BSc; Ramin Khatami[1,2,3], MD

[1]Center for Sleep Medicine, Sleep Research and Epileptology, Barmelweid, Switzerland

[2]Barmelweid Academy, Clinic Barmelweid AG, Barmelweid, Switzerland

[3]Department of Neurology, Bern University Hospital and University of Bern, Bern, Switzerland

Corresponding Author:
Zhongxing Zhang, PhD
Center for Sleep Medicine
Sleep Research and Epileptology
Clinic Barmelweid AG
Barmelweid, CH-5017
Switzerland
Phone: 41 62 857 22 38
Email: zhongxing.zhang@barmelweid.ch

Abstract

Background: Obstructive sleep apnea (OSA) is the most prevalent respiratory sleep disorder occurring in 9% to 38% of the general population. About 90% of patients with suspected OSA remain undiagnosed due to the lack of sleep laboratories or specialists and the high cost of gold-standard in-lab polysomnography diagnosis, leading to a decreased quality of life and increased health care burden in cardio- and cerebrovascular diseases. Wearable sleep trackers like smartwatches and armbands are booming, creating a hope for cost-efficient at-home OSA diagnosis and assessment of treatment (eg, continuous positive airway pressure [CPAP] therapy) effectiveness. However, such wearables are currently still not available and cannot be used to detect sleep hypopnea. Sleep hypopnea is defined by ≥30% drop in breathing and an at least 3% drop in peripheral capillary oxygen saturation (Spo_2) measured at the fingertip. Whether the conventional measures of oxygen desaturation (OD) at the fingertip and at the arm or wrist are identical is essentially unknown.

Objective: We aimed to compare event-by-event arm OD (arm_OD) with fingertip OD (finger_OD) in sleep hypopneas during both naïve sleep and CPAP therapy.

Methods: Thirty patients with OSA underwent an incremental, stepwise CPAP titration protocol during all-night in-lab video-polysomnography monitoring (ie, 1-h baseline sleep without CPAP followed by stepwise increments of 1 cmH_2O pressure per hour starting from 5 to 8 cmH_2O depending on the individual). Arm_OD of the left biceps muscle and finger_OD of the left index fingertip in sleep hypopneas were simultaneously measured by frequency-domain near-infrared spectroscopy and video-polysomnography photoplethysmography, respectively. Bland-Altman plots were used to illustrate the agreements between arm_OD and finger_OD during baseline sleep and under CPAP. We used *t* tests to determine whether these measurements significantly differed.

Results: In total, 534 obstructive apneas and 2185 hypopneas were recorded. Of the 2185 hypopneas, 668 (30.57%) were collected during baseline sleep and 1517 (69.43%), during CPAP sleep. The mean difference between finger_OD and arm_OD was 2.86% (95% CI 2.67%-3.06%, t_{667}=28.28; *P*<.001; 95% limits of agreement [LoA] –2.27%, 8.00%) during baseline sleep and 1.83% (95% CI 1.72%-1.94%, t_{1516}=31.99; *P*<.001; 95% LoA –2.54%, 6.19%) during CPAP. Using the standard criterion of 3% saturation drop, arm_OD only recognized 16.32% (109/668) and 14.90% (226/1517) of hypopneas at baseline and during CPAP, respectively.

Conclusions: arm_OD is 2% to 3% lower than standard finger_OD in sleep hypopnea, probably because the measured arm_OD originates physiologically from arterioles, venules, and capillaries; thus, the venous blood adversely affects its value. Our findings demonstrate that the standard criterion of ≥3% OD drop at the arm or wrist is not suitable to define hypopnea because it could provide large false-negative results in diagnosing OSA and assessing CPAP treatment effectiveness.

(J Med Internet Res 2021;23(7):e24171) doi:10.2196/24171

XSL•FO
RenderX

KEYWORDS

obstructive sleep apnea; wearable devices; smartwatch; oxygen saturation; near-infrared spectroscopy; continuous positive airway pressure therapy; photoplethysmography

Introduction

Monitoring health using wearables, such as smartwatches and armbands, is becoming a new lifestyle [1-4]. Hundreds of millions of smartwatches and armbands are being used daily, and the number is still sharply increasing. Sleep monitoring is one of the most popular functions of such wearables [2,5-7], because sleep is a critical determinant of an individual's health and well-being. Obstructive sleep apnea (OSA) is the most prevalent respiratory sleep disorder occurring in 9% to 38% of the general population [8], and it is a high-risk factor for cardio- and cerebrovascular diseases [9,10]. Nevertheless, about 90% of suspected patients with OSA remain undiagnosed [11] due to the lack of sleep laboratories or specialists and the high cost associated with an in-lab polysomnography (PSG) diagnosis (ie, the gold-standard for sleep disorder diagnosis [12]), thus leading to decreased quality of life and increased health care burden in the aging society. Relatively simple and less-expensive diagnostic methods such as portable home respiratory polygraphy have been developed [13,14]. However, its signal quality is compromised, with failure rates ranging from 3% to 18% [14], mostly due to incorrect handling of the device or sensors by the people performing the test (ie, the patients themselves) [15]. Home respiratory polygraphy also has the risk of false diagnosis because it cannot measure sleep/wake, so patients may be awake during the night. The cost of such diagnostic methods is still relatively high in many low- and middle-income countries, limiting their broad application in the general population. Therefore, using low-cost and easy-to-use wearable devices, such as smart finger rings, smartwatches, or armbands, for at-home diagnosis of OSA and to assess treatment effectiveness would substantially contribute to public health worldwide [6,16].

However, using the aforementioned low-cost wearable devices to evaluate OSA is still not clinically viable because, currently, no product has been licensed or certified as a medical diagnostic device by the United States Food and Drug Administration (FDA) or CE marking. We hypothesize that one of the major limitations in measuring OSA with wearables is the detection of sleep hypopnea. Currently, most consumer-grade wearables can measure surrogate markers for breathing and heartbeats or heart rate variability [3,17-20]. Although sleep apneas consisting of a complete pause in breathing for ≥10 seconds are relatively easy to assess by analyzing breathing frequency [6,12,21-23], it is challenging to detect sleep hypopnea, which is defined as a ≥30% drop in airflow lasting ≥10 seconds accompanied by either an arousal or a ≥3% drop in peripheral capillary oxygen saturation (SpO_2) measured at the fingertip [12]. Smart rings that can measure SpO_2 at the fingertip (eg, Sleepon [24]) are likely to accurately quantify the drops in SpO_2 because, essentially, they are similar to fingertip pulse oximetry. However, these devices have only a very tiny market share compared to other popular wearables such as smartwatches and armbands [25]. Whether the measures of oxygen desaturation (OD) at the fingertip and at the arm or wrist are physiologically identical in sleep hypopneas is essentially unknown. Some recent smartwatches (eg, Fitbit [26], Garmin [27], Huami [28], and Huawei [29]), armbands (eg, Humon [30,31], Moxy [32,33], PortaMon [34], and Biofourmis [35]), and prototypes [36,37] claim to measure SpO_2 or muscle tissue oxygen saturation (StO_2) at the arm or wrist. However, to the best of our knowledge, these devices have not been clinically validated for use in patients with OSA. We were able to find only one registered clinical validation study (Trial Registration: ClinicalTrials.gov NCT03775291) measuring OSA using a smartwatch, which was initiated by Fitbit in December 2018 [38]. However, the latest update of the study's recruitment status as of November 2019 was still "active, not recruiting." The study aimed to only compare PSG-assessed Apnea–Hypopnea Index (AHI) with the AHI derived from the smartwatch, rather than performing an event-by-event comparison of the apnea or hypopnea events measured by these two devices. Thus, even if we assume that Fitbit may have completed their validation work recently, their study still cannot answer the key question as to whether the hypopnea diagnostic standard criterion of ≥3% OD at the fingertip can be equally applied to the OD at the arm or wrist.

Therefore, we, for the first time, aimed to compare event-by-event OD at the fingertip (finger_OD) measured by the gold-standard in-lab PSG transmission photoplethysmography (T-PPG) with OD at the arm (arm_OD) measured by frequency-domain multidistance (FDMD) near-infrared spectroscopy (NIRS) in sleep hypopneas during naïve sleep and continuous positive airway pressure (CPAP) therapy. FDMD-NIRS is a well-validated [39-41] advanced, noninvasive optical technique that can quantify hemodynamic changes, including OD, in the measured tissues for long-term recordings such as all-night sleep with high temporal resolution [42,43]. Our results can conclusively demonstrate whether physiologically arm_OD measures can directly replace finger_OD to define sleep hypopnea. Thus, this study will have a broad appeal to the general public, sleep clinicians and scientists, health care insurance providers, and wearable technology developers who are aiming to measure OSA at-home by using wearable devices.

Methods

Patients

All patients underwent video-PSG measurement for diagnosis in our sleep laboratory. The following day, those patients who were diagnosed with OSA and recommended to use CPAP therapy by clinicians were recruited and gave their written informed consent for participation in the study. Patients with unstable coronary or cerebral artery disease, severe arterial hyper- or hypotension, respiratory diseases, or a history of a sleep-related accident were excluded. Finally, 30 newly diagnosed patients with OSA (mean age 54.2, SD 13.8 years, IQR 42-65 years; male: n=27; mean BMI 35.9, SD 7.5 kg/m^2,

XSL•FO
RenderX

IQR 31.8-42.0 kg/m^2; mean AHI 53.4, SD 24.7 per hour, IQR 32-71 per hour) participated in this study. This study was approved by the local ethical commission of Northwest Switzerland, and it was in compliance with the Declaration of Helsinki.

Protocol

The patients underwent incremental stepwise CPAP (AirSense 10, ResMed) titration combined with video-PSG and FDMD-NIRS recordings in one nocturnal sleep episode. This sleep episode consisted of 1 hour of baseline sleep without CPAP, followed by incremental stepwise titration of 1 cmH_2O pressure per hour starting from 5-8 cmH_2O depending on the individuals. The 1-h baseline sleep allowed us to compare arm_OD with finger_OD during natural sleep in patients with OSA. We included the CPAP titration protocol because (1) the stepwise CPAP titration protocol can increase the number of hypopneas for data analysis in our patients, since low pressures cannot fully open the airways to restore the apneas but instead cause hypopneas and (2) auto-CPAP that can automatically adjust the pressures within a given range (ie, automatic titration), and CPAP with fixed pressures are the most efficient and popular therapy for OSA [44]. Thus, the comparison between arm_OD and finger_OD in hypopneas under various CPAP pressures could allow us to test the feasibility of monitoring treatment efficacy by measuring arm_OD.

Video-PSG

Video-PSG (Embla RemLogic, Natus Medical Incorporated) is a comprehensive recording of physiological signals during sleep, including electroencephalography at electrode locations of C3, C4, O1, O2, F3, and F4 according to 10 to 20 system, eye movements (electrooculogram), muscle activation (electromyogram), electrocardiogram, breathing functions, heart rate, fingertip SpO_2 (left index fingertip in this study), and movement during sleep. Two experienced sleep technologists independently scored the sleep stages, respiratory and limb movement events, and motion artifacts in 30-second epochs according to the 2017 American Academy of Sleep Medicine manual [12]. Sleep hypopneas were defined as an at least 30% drop in airflow lasting for at least 10 seconds with either an arousal or >3% drop in SpO_2. The discrepancy between these two technologists was resolved by discussion or recommendation by an experienced neurophysiologist. The hypopneas were excluded from analysis if their SpO_2 desaturations were larger than 15% (n=31) to exclude outliers and potentially unreliable measurements caused by instrument errors.

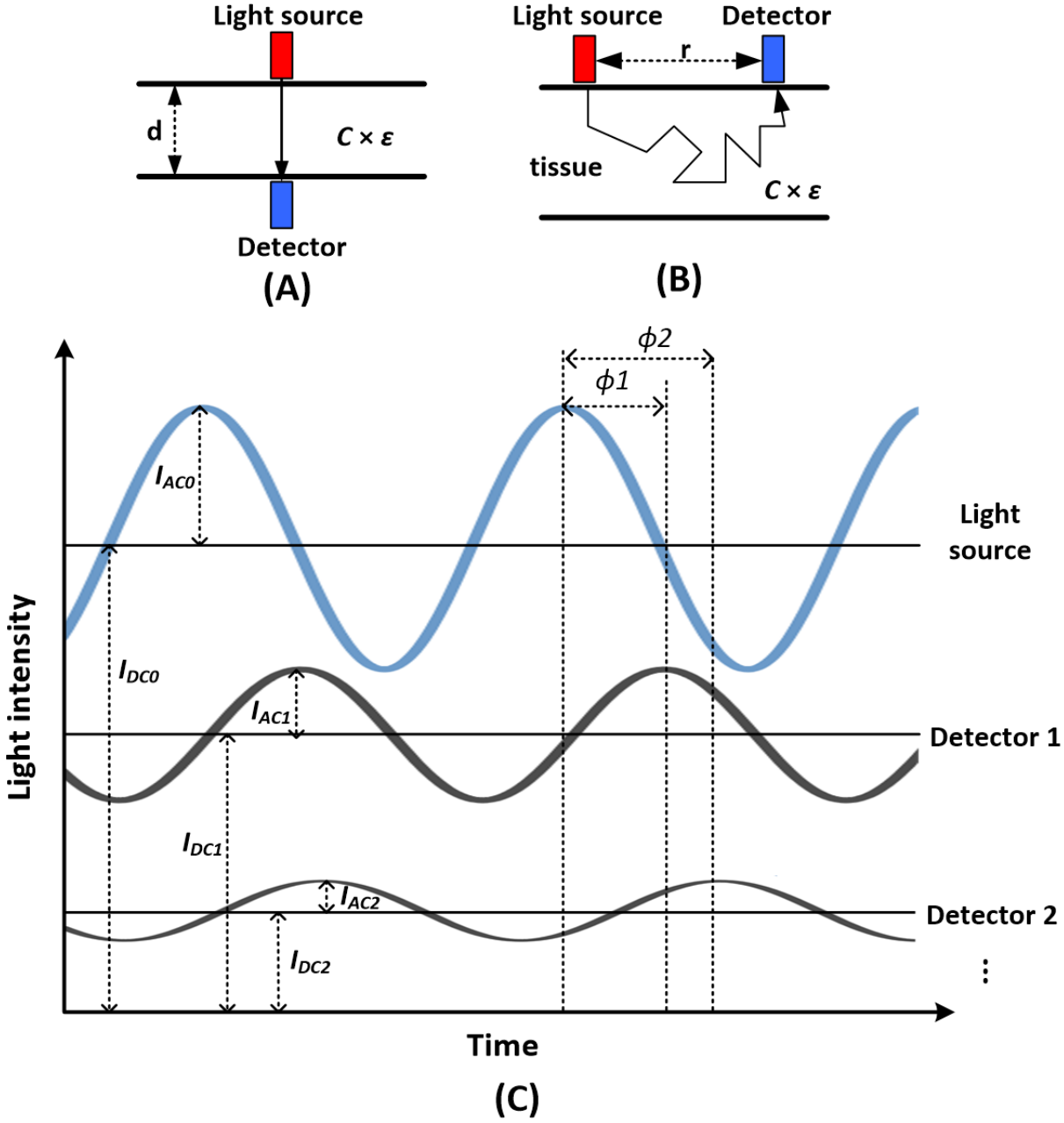
FDMD-NIRS Measurements

In this study, FDMD-NIRS (Imagent, ISS) measurements were conducted over the middle of the left biceps muscle. Imagent is currently the only commercial benchtop FDMD-NIRS device [40,42,45] and has been CE-approved for research. Its light emitters, four laser diodes at 690-nm wavelength and four laser diodes at 830-nm wavelength, are coupled into four light sources and are high frequency modulated at 110 MHz. The light can penetrate the measured tissues with a depth of several centimeters when the four light sources are aligned and placed at 2, 2.5, 3, and 3.5 cm away from an optical fiber bundle connected to the photomultiplier tube detector.

The most common commercially available NIRS devices, including wearable NIRS devices (eg, Humon [30,31], Moxy [32,33], and PortaMon [34]), are continuous-wave NIRS (CW-NIRS), which measure the hemodynamic changes in human tissues based on the modified Beer–Lambert law (MBLL) [46-48]. As illustrated in Figure 1A, in the original Beer–Lambert law, the light extinction is proportional to the concentration C multiplied by the constant extinction coefficient ε for the particular absorber and the length d of the absorbing media when light passes through a nonscattering but absorbing media [49]. $C \times \varepsilon$ is also called the absorption coefficient μ_a of the absorbing media. However, biological tissues are highly scattering media, and scattering will increase the path-length of light, thus increasing the probability of both light absorption and loss of light. Scattering also makes it possible that some light can go out of the tissue from the same side of the light source (ie, backscattering light), as shown in Figure 1B. Thus, reflectance photoplethysmography (R-PPG) pulse oximeter [50-52] and the NIRS device in which the light source and detector are placed on the same side of the measured tissues were developed based on MBLL. In MBLL, the light propagation due to scattering is taken into account by introducing the differential path-length factor (DPF). The real path-length of light in the tissue is then calculated as DPF multiplied by the source-detector distance r. DPF varies between different biological tissues and different individuals, and it also depends on other factors such as light wavelength and age and gender of the individual [48,53]. CW-NIRS devices use fixed value of DPF in the range of 3 to 6 for different light wavelengths [53]. They can only estimate the relative changes in the main absorbing chromophores, that is, oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HHb) in the measured biological tissues. R-PPG and T-PPG pulse oximeters usually use two near-infrared wavelengths that are mainly absorbed by HbO_2 and HHb respectively. The influences of scattering on light attenuation are approximately assumed to be cancelled out in the calculation of SpO_2 that is derived from the ratio of the light intensities of the two wavelengths, ie, assuming the path-lengths of the two wavelengths are identical in the tissues [50-52]. However, this assumption is actually not valid because DPF varies between different wavelengths. Experimental calibrations against in vitro measurement of arterial oxygen saturation (SaO_2) in extracted arterial blood (ie, invasive co-oximetry) thus must be performed in commercial pulse oximeters during their research and development period to correct measurement errors [50-52]. Several calibration-free methods were also proposed, including calculating the real light path-length and absorption by using the frequency-domain NIRS algorithm [52,54-57].

XSL•FO
RenderX

Figure 1. Beer–Lambert law and frequency-domain multidistance near-infrared spectroscopy (FDMD-NIRS) measurement. (A) The original Beer–lambert law describing the light propagation in a nonscattering absorbing media. The attenuation of light intensity in this absorbing media is proportional to the concentration C multiplied by the constant extinction coefficient ε for the particular absorber and the length d of the absorbing media. (B) The travelling of light in biological tissue (ie, highly scattering media). Light travels a longer pathway in the tissue than the light source-detector distance r due to scattering. (C) Basic principle of FDMD-NIRS measurement. The blue sine wave represents the high-frequency modulated light source. I_{DC0} and I_{AC0} are its light intensity and modulation amplitude, respectively. The two black sine waves are the output light detected after passing the measured tissues. They are detected by detectors 1 and 2 placed at different distances away from the light source. The light intensities and modulation amplitudes of the two black sine waves are smaller than those of the light source, and their phases are delayed because of the absorption and scattering in the tissues. Detector 1 is closer to the light source than detector 2. Therefore, the light intensity I_{DC1} and modulation amplitude I_{AC1} detected at detector 1 are larger than the light intensity I_{DC2} and modulation amplitude I_{AC2} detected at detector 2. The phase delay ϕ1 at detector 1 is smaller than the phase delay ϕ2 at detector 2 because the light reaching detector 2 has travelled a longer distance in the tissue. Similarly, the light intensity and modulation amplitude will be further decreased, and the phase delay will be further increased, when the light reaches the other detectors placed farther away than detector 2.



As illustrated in Figure 1C, in FDMD-NIRS, the light emitted from light source can be detected by detectors placed at different distances away from the source. The light intensity (I_{DC}) and modulation amplitude (I_{AC}) of the detected light decrease, and a phase delay (ϕ) occurs between the detected light and the source light due to absorption and scattering. The detected I_{DC} and I_{AC} are smaller but theϕ is larger at the detector further away from the light source. The I_{DC}, I_{AC} and ϕ from different light source-detector distances vary linearly [42,45]. Therefore, to submit the measured I_{DC}, I_{AC} and ϕ to linear regression, we can obtain the following equations derived from the photon diffusion equation in a semi-infinite geometry [40,45,58-60]:

$$ln\,(r^2 I_{AC}) = rS_{AC} + C_{AC} \quad \textbf{(1)}$$

XSL•FO
RenderX

$$ln\,(r^2 I_{DC}) = rS_{DC} + C_{DC} \quad \textbf{(2)}$$

$$\phi = rS_{\phi} + C_{\phi} \quad \textbf{(3)}$$

Where r is the known source-detector distance, S_{AC}, S_{DC} and S_{ϕ} are the slopes and C_{AC}, C_{DC} and C_{ϕ} are the intercepts. The linearity of the equation is monitored by the R^2 of the fitted linear regression. By combing any two of these three slopes (eg, we chose S_{AC} and S_{ϕ} in the following equations), we can calculate the absorption coefficient μ_a and the reduced scattering coefficient μ_S' of the measured tissue [40,45,58-60]:

$$\mu_a = \omega/2v \times (S_{\phi} / S_{AC} - S_{AC} / S_{\phi}) \quad \textbf{(4)}$$

$$\mu_S' = (S_{AC}^2 - S_{\phi}^2) / 3\mu_a - \mu_a \quad \textbf{(5)}$$

where $\omega/2\pi$ is the modulation frequency and v is the velocity of light in the tissue. FDMD-NIRS uses two wavelengths. The μ_a and μ_S' of both these wavelengths can be calculated individually by using the same equations (4) and (5). Equation (4) gives us the absorption coefficient of the measured tissues calculated by taking the influence of scattering into account. It is equal to $C \times \varepsilon$ as mentioned above. In NIRS, the main contributions to absorptions in tissues are HbO_2 and HHb, so we have the following equation:

$$\mu_a^{\lambda} = \varepsilon_{HHb}^{\lambda} C_{HHb} + \varepsilon_{HbO2}^{\lambda} C_{HbO2} \quad \textbf{(6)}$$

where μ_a^{λ} is the absorption coefficient of the measured tissue at wavelength λ. $\varepsilon_{HHb}^{\lambda}$ and $\varepsilon_{HbO2}^{\lambda}$ are the known extinction coefficients at wavelength λ for HHb and HbO_2, respectively. C_{HHb} and C_{HbO2} are the concentrations of HHb and HbO_2, respectively. Using two wavelengths $\lambda 1$ and $\lambda 2$, we can then calculate C_{HHb} and C_{HbO2} with the following equations:

$$C_{HbO2} = (\mu_a^{\lambda 1} \varepsilon_{HHb}^{\lambda 2} - \mu_a^{\lambda 2} \varepsilon_{HHb}^{\lambda 1}) / (\varepsilon_{HbO2}^{\lambda 1} \varepsilon_{HHb}^{\lambda 2} - \varepsilon_{HbO2}^{\lambda 2} \varepsilon_{HHb}^{\lambda}) \quad \textbf{(7)}$$

$$C_{HHb} = (\mu_a^{\lambda 2} \varepsilon_{HbO2}^{\lambda 1} - \mu_a^{\lambda 1} \varepsilon_{HbO2}^{\lambda 2}) / (\varepsilon_{HbO2}^{\lambda 1} \varepsilon_{HHb}^{\lambda 2} - \varepsilon_{HbO2}^{\lambda 2} \varepsilon_{HHb}^{\lambda 1}) \quad \textbf{(8)}$$

Therefore, Sto_2 can be further derived as:

$$StO_2 = 100 \times C_{HbO2} / (C_{HbO2} + C_{HHb}) \quad \textbf{(9)}$$

The FDMD-NIRS algorithm can calculate the absolute values of HbO_2, HHb, and StO_2 in the measured tissue and it is superior to the simple CW-NIRS algorithm, because of its sophisticated mathematical frameworks calculating μ_a and μ_S' that can best estimate the real light propagation distance in the measured tissues based on the diffusion equation in complex geometries. The robustness, precision, and accuracy of measuring HbO_2, HHb, and StO_2 of the Imagent system used in this study have been well validated in different physical blood-lipid models [45,58,61] and in vivo studies [59,62-64]. It has been used as a gold-standard reference measurement of StO_2 for validations or calibrations of wearable CW-NIRS armbands [31] and portable CW-NIRS oximeters [61,65] including those that have received FDA clearance [66,67].

In this study, the sampling rate of FDMD-NIRS recording was 5.2 Hz. The reliability and accuracy of FDMD-NIRS measurements depend on the linearity of the measured optical signals on distances, because μ_a and μ_S' are derived from the slopes of equations (1-3). The linear dependence R^2 of the modulated light amplitude and phase shift over the measured distances should be highly close to 1 in each light wavelength. Thus, before the start of the recording, our Imagent system was calibrated on an optical phantom block with known μ_a and μ_S', that is, the light intensity of each light source was adjusted so that R^2 was equal to 1 and the measured μ_a and μ_S' of the optical phantom block were equal to their known values. This calibration step can exclude the uncertainty of our measurements due to machine errors such as light source and detector errors. The raw optical data were discarded if the R^2 was smaller than 0.95 in either modulated light amplitude or phase shift in any wavelength to exclude poor-quality data arising from improper probe-skin contact and shunted light reaching the detector without travelling through the tissue [68,69]. The NIRS data were then subjected to a low-pass (<0.08 Hz), zero-phase filter designed using Hanning window to remove the physiological noises, including heart rate, respiratory noise, and spontaneous slow hemodynamic oscillations [70,71]. The filtered data were smoothed with moving average smooth method (robust locally weighted scatter plot smoothing [70,72]).

Statistical Analysis

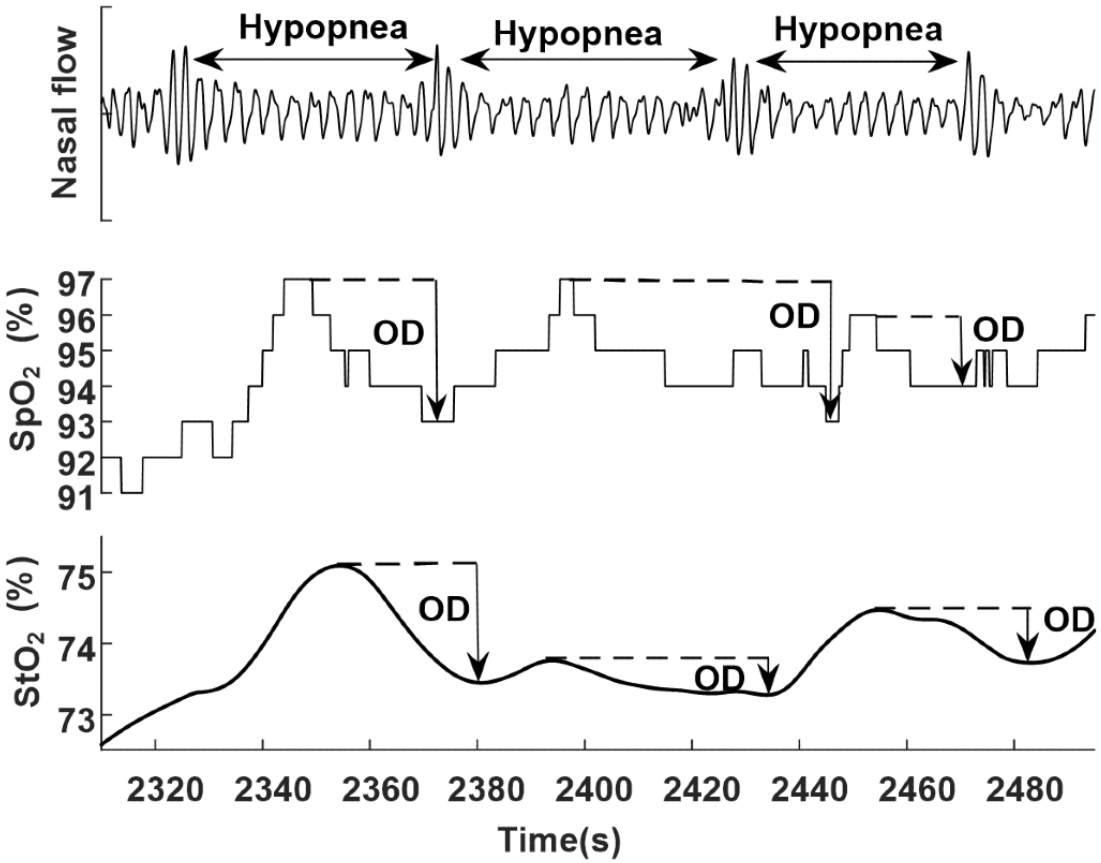
Bland-Altman plots were used to illustrate the agreements between arm_OD and finger_OD during baseline sleep and under CPAP, respectively. We used *t* test to determine whether the differences between arm_OD and finger_OD were significantly different from 0. In order to check whether arm_OD could replace finger_OD in subgroups of hypopneas (ie, hypopneas with different degrees of desaturations), we used Spearman correlation to evaluate the relationship between arm_OD and finger_OD in different subgroups defined by finger_OD greater than or equal to specific cut-offs. The cut-offs ranged from 2% to 8%. *P* values <.05 indicated statistical significance for both analyses. Pre-processing of FDMD-NIRS signals was carried out in MATLAB (MathWorks, Inc.). All statistical analyses were performed using R statistical software (version 3.2.4; R Foundation for Statistical Computing).

Results

In total, 2185 hypopneas (median 67, IQR 41-82) were analyzed, including 668 (median 16, IQR 11-35) recorded during baseline sleep and 1517 (median 41, IQR 25-64) recorded during CPAP. Figure 2 illustrates the typical changes that occurred in arm_OD and finger_OD in hypopnea events. Indeed, sleep hypopneas cause OD in arm StO_2. The absolute (mean) values of StO_2 are 68.8% (SD 6.5%) at baseline before the start of hypopneas. The distributions of finger_OD and arm_OD (Figure 3) suggest larger OD in the finger than in the arm both at the baseline and during CPAP, as the peaks of the distributions of arm_OD are smaller than those of finger_OD.

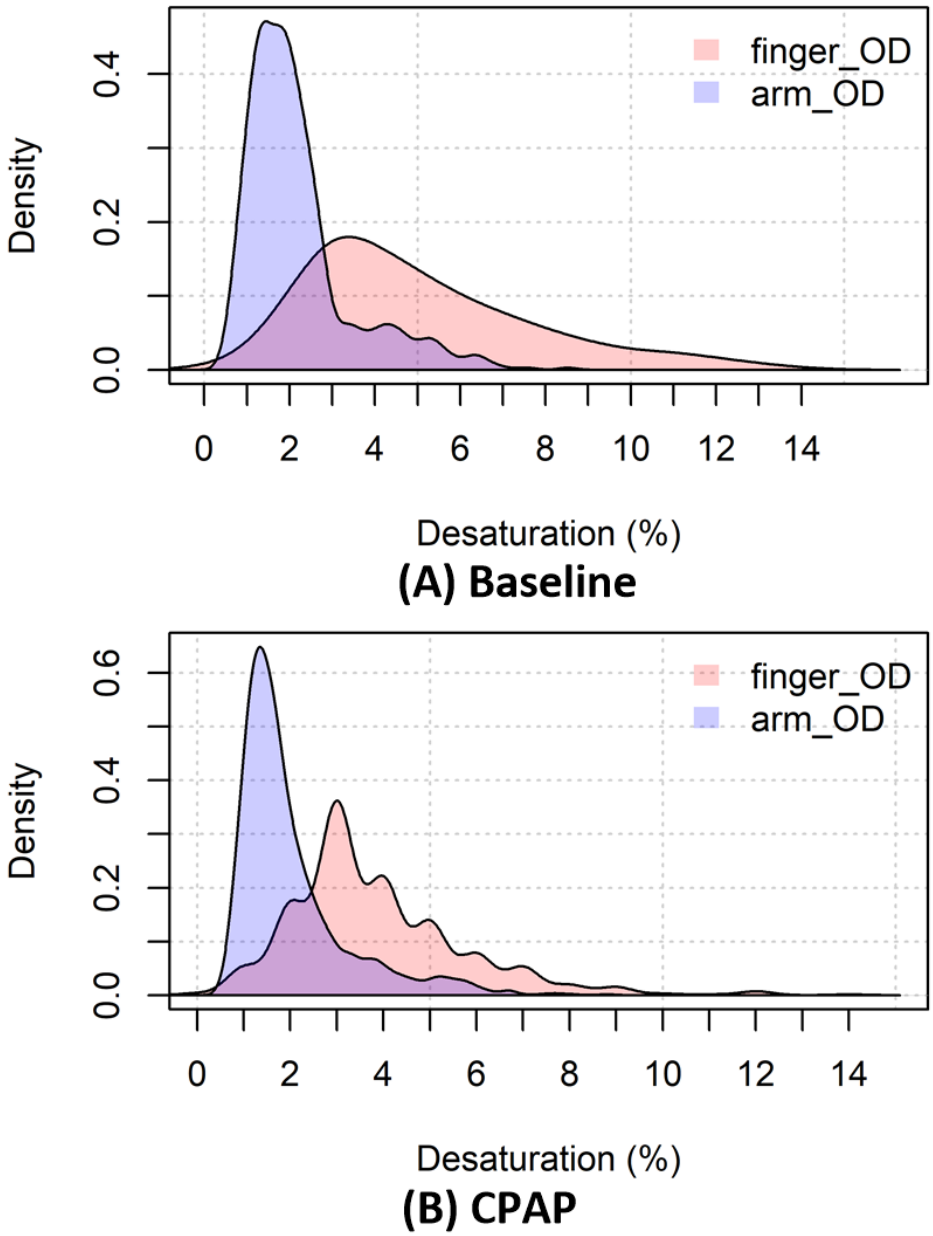
XSL•FO
RenderX

Figure 2. Typical oxygen desaturation (OD) at fingertip (finger_OD) and at arm (arm_OD) during hypopneas. Arrows indicate the degree of OD. SpO_2 is measured at the fingertip by polysomnography transmission photoplethysmography, and StO_2 is measured at the biceps muscle by frequency-domain multidistance near-infrared spectroscopy. SpO_2: peripheral capillary oxygen saturation; StO_2: peripheral tissue oxygen saturation.



XSL•FO
RenderX

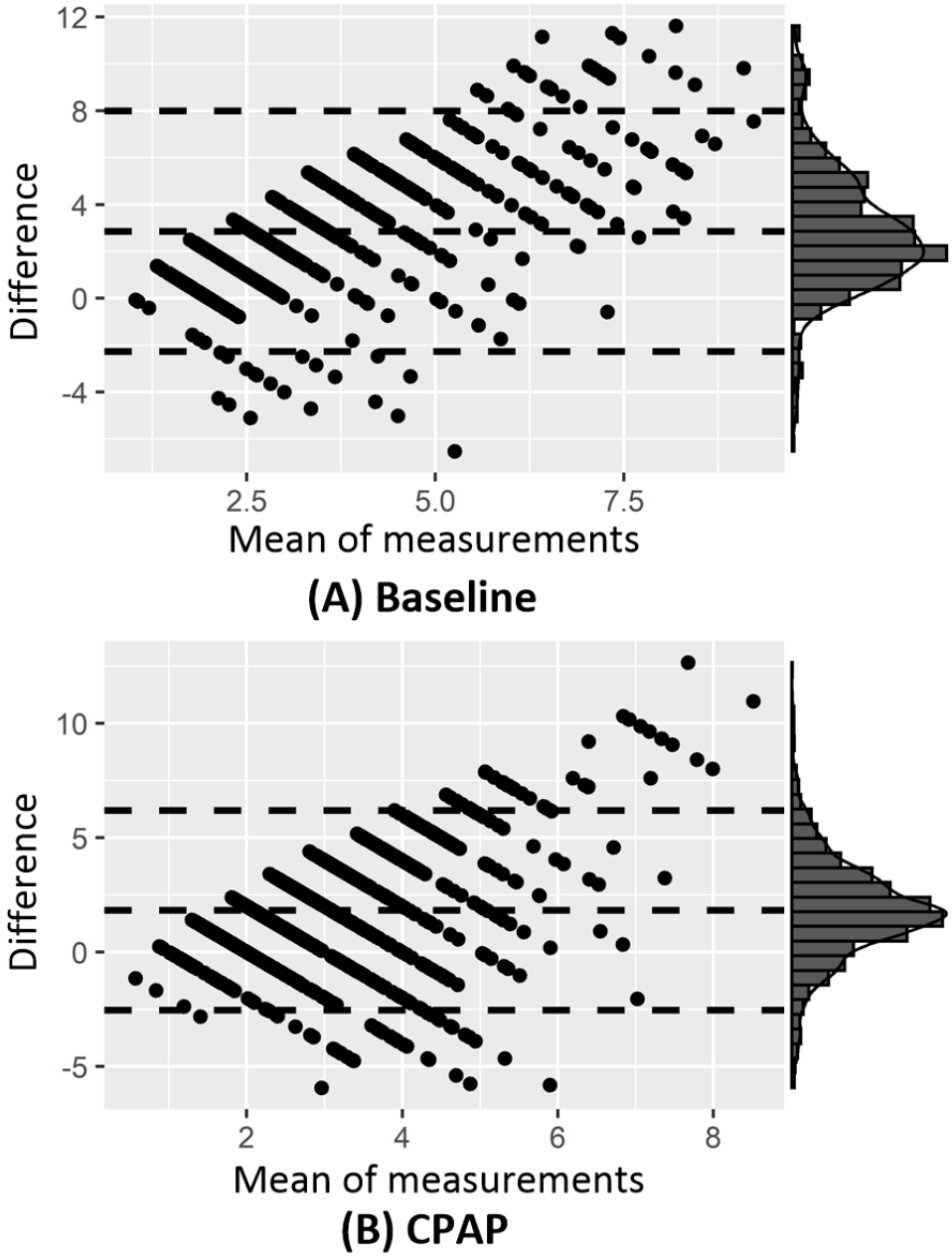
Figure 3. Distributions of oxygen desaturation at fingertip (finger_OD) and at arm (arm_OD) at (A) baseline (n=668) and (B) under continuous positive airway pressure (CPAP: n=1517).



Bland-Altman plots (Figure 4) show that the mean difference between finger_OD and arm_OD is 2.86% (95% CI 2.67%-3.06%, t_{667}=28.28; P<.001) during baseline sleep and 1.83% (95% CI 1.72%-1.94%, t_{1516}=31.99; P<.001) under continuous positive airway pressure (CPAP) sleep, with broad 95% limits of agreement (LoA) as [-2.27%, 8.00%] and [-2.54%, 6.19%], respectively. Using the criterion of arm_OD ≥3%, we can only define 16.32% (109/668) and 14.90% (226/1517) of hypopneas at baseline and during CPAP sleep, respectively.

XSL•FO
RenderX

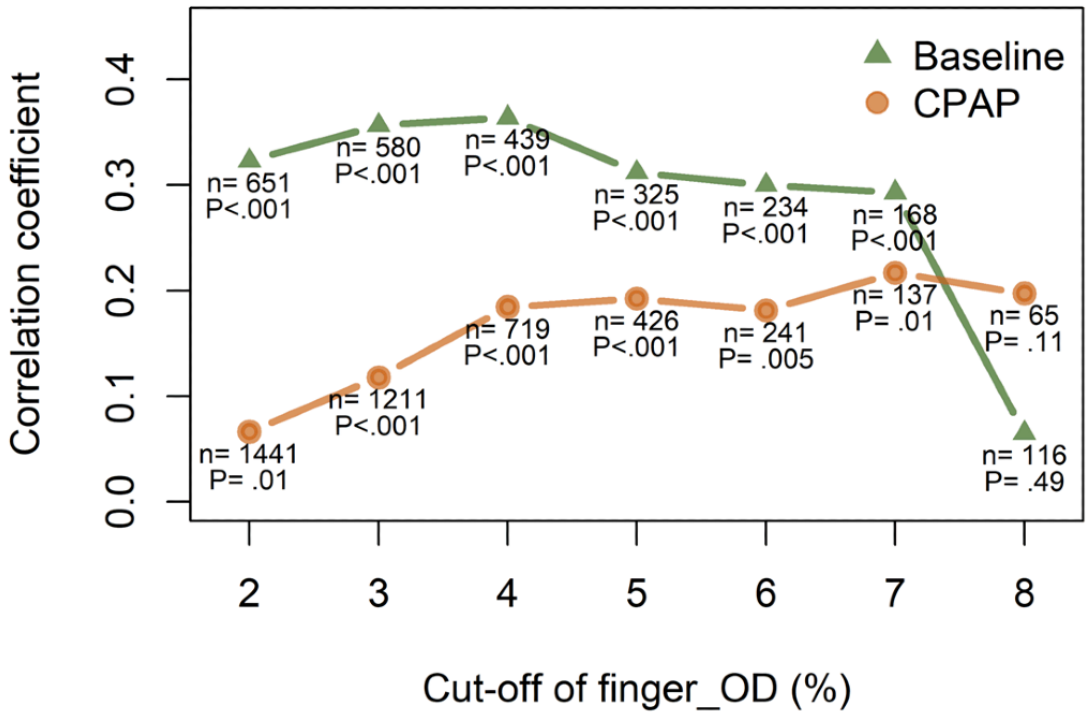
Figure 4. Bland-Altman plots of oxygen desaturation at fingertip (finger_OD) and at arm (arm_OD) at (A) baseline (n=668) and (B) under continuous positive airway pressure (CPAP: n=1517). The x-axes show the mean between the two measures, whereas the y-axes represent the differences (ie, finger_OD – arm_OD). The horizontal dotted lines indicate the mean difference and the 95% limits of agreement between the measures, ie, mean difference ± 1.96 × SD. The distribution of the mean difference is shown at the right margin of the plot, which is a normal distribution.



To test whether arm_OD could replace finger_OD in subgroups of hypopneas such as those causing severe OD, we correlate arm_OD and finger_OD in different subgroups with finger_OD ≥ specific cut-offs (from 2% to 8%). Again, we could only observe weak correlations (ie, correlation coefficients < 0.4) between them at baseline and under CPAP (Figure 5).

XSL•FO
RenderX

Figure 5. The correlations between oxygen desaturation at fingertip (finger_OD) and at arm (arm_OD) in the hypopnea events wherein finger_OD ≥ cut-off. X-axis shows the cut-off of finger_OD (from 2% to 8%). Y-axis depicts nonparametric Spearman's correlation coefficient between finger_OD and arm_OD. The number of events used for correlation analysis and the *P* value are shown in the figure. The correlation analyses are performed for both baseline and CPAP sleep. For example, the green triangle at X=6, Y=0.3 means that during naïve baseline sleep there are 234 hypopneas associated with at least 6% finger_OD, and in these events, arm_OD weakly correlates to finger_OD with Spearman's correlation coefficient equal to 0.3 and *P*<.001.



Since the baseline values of StO_2 (mean 68.8% SD 6.5%) are obviously much smaller than those of SpO_2 (usually above 90%), it is possible that the relative change rather than the absolute (raw) value of arm_OD has better agreement with the finger_OD. We therefore normalize the arm_OD to its baseline. The mean difference between finger_OD and normalized arm_OD is 1.68% (95% CI 1.46-1.90%, t_{667}=14.99; *P*<.001) during baseline sleep and 0.88% (95% CI: 0.75-1.01%, t_{1516}=12.91; *P*<.001) during CPAP sleep, respectively. The mean differences are smaller than those between absolute arm_OD and finger_OD. However, the 95% LoA are still equally broad, which are [-3.99%, 7.34%] during baseline sleep and [-4.32%, 6.08%] during CPAP sleep respectively, suggesting poor agreement. Only 41.6% (278/668) of hypopneas at baseline and 31.7% (481/1517) during CPAP sleep can be defined using the criterion of normalized arm_OD ≥3%, suggesting that the normalization still has poor sensitivity in detecting hypopneas.

Discussion

Principal Findings

In this study, we compare, for the first time, event-by-event finger_OD to arm_OD in sleep hypopneas during naïve sleep and CPAP therapy sleep. We choose the gold-standard reference methods for both arm_OD and finger_OD—advanced FDMD-NIRS for arm_OD and PSG fingertip transmission photoplethysmography for finger-OD. Arm_OD is 2% to 3% smaller than finger_OD, probably because finger_OD is caused only by arterioles, whereas arm_OD is physiologically determined by mixed sources of arterioles, venules, and capillaries [42,73,74]. The lower value of arm_OD is thus most likely due to the contribution of the venous blood pool. The significant difference between finger_OD and arm_OD and their broad LoA (Figure 4) suggest that arm_OD cannot directly replace finger_OD to define hypopneas. If the standard criterion of ≥3% drop is applied, it will cause a high rate of false-negative results in diagnosing OSA and in assessing the efficacy of CPAP treatment. The poor agreement between finger_OD and arm_OD and the low sensitivity of arm_OD in measuring hypopneas cannot essentially be improved even after normalization, in which raw arm_OD is normalized to its baseline before the start of hypopnea. The poor correlations (Figure 5) between finger_OD and arm_OD across hypopnea severity further suggest that the arm_OD cannot be used to define hypopneas, that is, correlations remain low even in severe hypopneas that are associated with much larger finger_ODs.

The armband and smartwatch are two of the most popular wearable technologies. Our results have direct implications for wearable armbands and portable oximeters using NIRS techniques to measure OSA, as we discuss below.

First, the FDMD-NIRS system used in our study is well recognized as the most robust and reliable reference NIRS technique [31,42,61]. Arm StO_2 measured by our device should be identical to that measured by consumer-grade wearable NIRS armbands (eg, Humon [30,31], Moxy [32,33], and PortaMon [34]). It should also be equal to that measured by FDA-cleared medical-grade portable NIRS oximeters, such as INVOS 5100C (Medtronic) [66], FORE-SIGHT (CAS Medical Systems) [67], SenSmart (Nonin Medical Inc) [75], Hutchinson InSpectra

XSL•FO
RenderX

(Hutchinson Technology Inc) [76], and ViOptix ODISsey (ViOptix Inc) [77]. Our results suggest that the absolute or relative (ie, normalized to baseline) OD at arm muscle measured by these aforementioned devices cannot be directly used to define sleep hypopnea.

Second, the NIRS StO_2 is the proportion of HbO_2 in arterial, capillary, and venous compartments of the measured tissue. It can be expressed by its two major compounds as $StO_2 = (a \times SaO_2) + (b \times SvO_2)$, where SaO_2 and SvO_2 are the arterial and venous oxygen saturation [68,73]. Fingertip SpO_2 is the best noninvasive estimate of SaO_2 [50,52]. The ratio of the coefficients a/b is called the arterio-venous ratio (AVR). The AVR of different tissues (eg, brain and muscles) and different populations (eg, healthy people and patients, adults, and children) has been determined by invasive measurements under different conditions but not during sleep [73,78-88]. Commercially available NIRS oximeters, including the aforementioned FDA-certificated medical devices take the fixed AVR value as either 0.25/0.75 or 0.30/0.70 but have not validated it in OSA [78-88]. Our results of poor correlations between arm_OD and finger_OD (Figure 5) indicate that such an uncritical AVR adoption is not justified. According to the mathematical model of fixed relationship between StO_2 and SaO_2, we expect a strong correlation between StO_2 and SpO_2 that we could not confirm using our data. These conflicting results suggest that the model is not valid for OSA. Therefore, developers of wearable or portable NIRS devices should first study the AVR in muscle tissues and validate it against invasive blood sample measurements before using their devices commercially for OSA diagnosis.

Our findings also indicate that appropriate arterial oxygen saturation measurement is not possible with armbands using the R-PPG method. As mentioned in the Methods section, R-PPG essentially has the same theoretical limitations as CW-NIRS, such that the scattering of light in the human tissues cannot be calculated. R-PPG and T-PPG pulse oximeters estimate SpO_2 under the assumption that changes in blood volume only occur in arterial but not in the venous compartment. SpO_2 is measured from the fingertip or the superficial forehead because these locations are well perfused by arteries [36,52,89]. It is not recommended to measure SpO_2 by using R-PPG at the wrist or the arm, because of its low signal-to-noise ratio. Compared to the more precise fingertip T-PPG method, the low signal-to-noise ratio of R-PPG at the wrist or arm is about 10 times weaker due to various factors such as relative low blood perfusion and sensitivity to pressure and ambient light sources [51,89]. Additionally, the key assumption of constant venous blood volume is no longer valid at the wrist or arm [89,90]. Although recently, some smartwatches [26-29] and a few armbands [35-37] claimed that they can measure SpO_2 at the arm or wrist by using R-PPG, they actually measure both arterial and venous blood [89,90] similar to NIRS. A main difference between NIRS and R-PPG is the measurement depth. Whether R-PPG measures the blood only in the skin or in both the skin and muscle depends on the distance between the light source and detector [91]. The detector can detect the light passing through deeper tissues at a larger separation distance. In vivo studies suggest that R-PPG can obtain its best signal-to-noise ratio at a separation distance of 3 to 6 mm [92]. Thus, a separation distance of several millimeters is used in the design of R-PPG pulse oximetry [36]. A recent study modeled the R-PPG light propagation in human skin [37]. The authors found that even at a separation distance of only 0.6 mm, many light rays reaching deeper into the muscle can still be received by the detector because of the random nature of light scattering [37]. Therefore, similar to NIRS armbands, armband devices using R-PPG capture the oxygen saturation in both skin and muscle. There are challenges in using both techniques to differentiate SaO_2 and SvO_2 desaturations from the measured StO_2 desaturation to define sleep hypopneas.

We did not include a smartwatch in this study, although recently, leading smartwatch companies like Apple, Fitbit, Garmin, Huami, and Huawei have all added the function of measuring oxygen saturation in their products. This is because these commercially available products cannot or are unwilling to export their raw data for analyses, or their temporal resolutions are simply too low for an event-by-event comparison necessary for our study because usually these consumer-grade smartwatches upload their data to their cloud servers in minute resolution. Nevertheless, our results may have indirect implications for smartwatches. R-PPG smartwatch is different from the fingertip T-PPG in measuring SpO_2. Lee et al [89] found that the raw light signals measured by wrist R-PPG and fingertip T-PPG change differently during breath-holding, indicating different SpO_2 values are calculated by these two techniques. Abay et al [36] reported that wrist R-PPG results in lower SpO_2 values than fingertip T-PPG at rest, and during venous occlusion, fingertip T-PPG SpO_2 does not exhibit desaturation but wrist R-PPG SpO_2 drops similarly as the simultaneously measured NIRS StO_2 at the same arm. Their findings also indicate that although smartwatches measure the wrist and NIRS measures the arm muscle, the measured changes in oxygen saturation by these two techniques are likely to be the same. Thus, the venous blood influences in our NIRS StO_2 measurements are also likely to be observed in smartwatch measurements.

Conclusions

Our study warns consumers, health care insurance companies, and sleep clinicians and scientists to interpret the AHI provided by smartwatches and armbands with caution until those products are clinically and experimentally validated. An AHI >5/hour suggests the diagnosis of OSA [12]. Our results suggest that AHI is likely to be underestimated if using the criterion of arm_OD ≥3% to define hypopneas. Wearable technology developers who are validating their products can learn from this study and take into account the mismatch between the ODs measured by their products and by the gold-standard technique fingertip pulse oximetry. Developing new parameters (eg, estimated-oxygen-variation provided by Fitbit smartwatches [26]) or combining smartwatches with external fingertip T-PPG sensors [93] may be a more promising strategy to measure OSA. Nevertheless, validations of these new approaches are necessary before releasing them for clinical use. In addition, our finding

XSL•FO
RenderX

of a weak correlation between finger_OD and arm_OD indicates that (1) prediction of finger_OD using arm_OD may be possible but will require development and implementation of sophisticated data-mining, such as machine learning algorithms [90], and (2) as previous studies have validated NIRS oximeters as medical devices, protocols that quantify the arterial and venous contributions to the arm_OD are needed. Arm_OD can then be calibrated to indicate the changes in the arterial ODs at the arm or wrist.

Acknowledgments

This work was supported by Clinic Barmelweid Scientific Foundation. The data acquisition work was supported by the Research Fund of the Swiss Lung Association No. 2014-22.

Conflicts of Interest

None declared.

References

1. Depner C, Cheng P, Devine J, Khosla S, de Zambotti M, Robillard R, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. Sleep 2020 Feb 13;43(2):1-13 [FREE Full text] [doi: 10.1093/sleep/zsz254] [Medline: 31641776]
2. Zhang Z, Cajochen C, Khatami R. social jetlag and chronotypes in the chinese population: analysis of data recorded by wearable devices. J Med Internet Res 2019 May 11;21(6):e13482 [FREE Full text] [doi: 10.2196/13482] [Medline: 31199292]
3. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]
4. Wen D, Zhang X, Liu X, Lei J. evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. J Med Internet Res 2017 Mar 07;19(3):e68 [FREE Full text] [doi: 10.2196/jmir.6874] [Medline: 28270382]
5. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
6. Baron KG, Duffecy J, Berendsen MA, Cheung Mason I, Lattie EG, Manalo NC. Feeling validated yet? a scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. Sleep Med Rev 2018 Aug;40:151-159 [FREE Full text] [doi: 10.1016/j.smrv.2017.12.002] [Medline: 29395985]
7. Reimer U, Emmenegger S, Maier E, Zhang Z, Khatami R. Recognizing sleep stages with wearable sensors in everyday settings. : SciTePress; 2017 Presented at: Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health; 2017; Porto p. 172-179. [doi: 10.5220/0006346001720179]
8. Senaratna CV, Perret JL, Lodge CJ, Lowe AJ, Campbell BE, Matheson MC, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. Sleep Med Rev 2017 Aug;34:70-81. [doi: 10.1016/j.smrv.2016.07.002] [Medline: 27568340]
9. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 2005 Nov 10;353(19):2034-2041. [doi: 10.1056/nejmoa043104]
10. Somers VK. Sleep--a new cardiovascular frontier. N Engl J Med 2005 Nov 10;353(19):2070-2073. [doi: 10.1056/NEJMe058229] [Medline: 16282183]
11. Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep 1997 Sep;20(9):705-706. [doi: 10.1093/sleep/20.9.705] [Medline: 9406321]
12. Berry R, Brooks R, Gamaldo CE, Harding SM, Lloyd RM, Marcus CL, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Darien, IL: The American Academy of Sleep Medicine; Oct 1, 2017.
13. Masa JF, Corral J, Pereira R, Duran-Cantolla J, Cabello M, Hernandez-Blasco L, et al. Effectiveness of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. Thorax 2011 May 20;66(7):567-573. [doi: 10.1136/thx.2010.152272]
14. Collop NA, Anderson WM, Boehlecke B, Claman D, Goldberg R, Gottlieb DJ, Portable Monitoring Task Force of the American Academy of Sleep Medicine. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. Portable Monitoring Task Force of the American Academy of Sleep Medicine. J Clin Sleep Med 2007 Dec 15;3(7):737-747 [FREE Full text] [Medline: 18198809]
15. Tan H, Gozal D, Ramirez H, Bandla H, Kheirandish-Gozal L. Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea. Sleep 2014 Feb 01;37(2):255-260 [FREE Full text] [doi: 10.5665/sleep.3392] [Medline: 24497654]
16. Penzel T, Schöbel C, Fietze I. New technology to assess sleep apnea: wearables, smartphones, and accessories. F1000Res 2018;7:413 [FREE Full text] [doi: 10.12688/f1000research.13010.1] [Medline: 29707207]

XSL•FO
RenderX

17. Zhang H, Zhang J, Li H, Chen Y, Yang B, Guo Y, et al. validation of single centre pre-mobile atrial fibrillation apps for continuous monitoring of atrial fibrillation in a real-world setting: pilot cohort study. J Med Internet Res 2019 Dec 03;21(12):e14909 [FREE Full text] [doi: 10.2196/14909] [Medline: 31793887]
18. Fan Y, Li Y, Li J, Cheng W, Shan Z, Wang Y, et al. Diagnostic performance of a smart device with photoplethysmography technology for atrial fibrillation detection: pilot study (Pre-mAFA II Registry). JMIR Mhealth Uhealth 2019 Mar 05;7(3):e11437 [FREE Full text] [doi: 10.2196/11437] [Medline: 30835243]
19. Chen E, Jiang J, Su R, Gao M, Zhu S, Zhou J, et al. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. Heart Rhythm 2020 May;17(5 Pt B):847-853 [FREE Full text] [doi: 10.1016/j.hrthm.2020.01.034] [Medline: 32354449]
20. Zhang Z, Henzmann S, Hügli G, Qi M, Chen W, Lu C, et al. Validation of wearable sleep monitoring device based on cardiopulmonary coupling and accelerometer with comparison to polysomnography in adults. : European Journal of Neurology; 2018 Presented at: The 4th Congress of the European Academy of Neurology; 2018; Lisbon p. 385.
21. Papini GB, Fonseca P, Gilst MMV, Bergmans JW, Vullings R, Overeem S. Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population. Physiol Meas 2020 Jul 08;41(6):065010. [doi: 10.1088/1361-6579/ab9481] [Medline: 32428875]
22. Taffoni F, Rivera D, La Camera A, Nicolò A, Velasco JR, Massaroni C. A wearable system for real-time continuous monitoring of physical activity. J Healthc Eng 2018;2018:1878354 [FREE Full text] [doi: 10.1155/2018/1878354] [Medline: 29849993]
23. Aliverti A. Wearable technology: role in respiratory health and disease. Breathe (Sheff) 2017 Jun;13(2):e27-e36 [FREE Full text] [doi: 10.1183/20734735.008417] [Medline: 28966692]
24. Go2sleep SE. URL: https://www.sleepon.us/go2sleepse/ [accessed 2020-08-07]
25. Seneviratne S, Hu Y, Nguyen T, Lan G, Khalifa S, Thilakarathna K, et al. A survey of wearable devices and challenges. IEEE Commun Surv Tutorials 2017 Jul 26;19(4):2573-2620. [doi: 10.1109/comst.2017.2731979]
26. How do I track my estimated oxygen variation in the Fitbit app?. URL: https://help.fitbit.com/articles/en_US/Help_article/1876.htm [accessed 2020-08-07]
27. Top FAQs for the Pulse Ox Feature on Garmin Watches. URL: https://support.garmin.com/en-US/?faq=SK2Y9a9aBp5D6n4sXmPBG7 [accessed 2020-08-07]
28. Amazfit X Bow to the Future. URL: https://www.amazfit.com/en/amazfit-x.html [accessed 2020-08-07]
29. HUAWEI Watch 3. Huawei Global. URL: https://consumer.huawei.com/en/wearables/watch-3/ [accessed 2021-07-07]
30. Humon Muscle Oxygen Sensor. URL: https://humon.io/ [accessed 2020-08-07]
31. Farzam P, Starkweather Z, Franceschini MA. Validation of a novel wearable, wireless technology to estimate oxygen levels and lactate threshold power in the exercising muscle. Physiol Rep 2018 Apr 02;6(7):e13664. [doi: 10.14814/phy2.13664]
32. The Science Behind Moxy. URL: https://www.moxymonitor.com/wp-content/themes/moxymonitor/documents/Moxy_Scientific_Explanation_march2014.pdf [accessed 2020-08-07]
33. Feldmann A, Schmitz R, Erlacher D. Near-infrared spectroscopy-derived muscle oxygen saturation on a 0% to 100% scale: reliability and validity of the Moxy Monitor. J Biomed Opt 2019 Nov;24(11):1-11 [FREE Full text] [doi: 10.1117/1.JBO.24.11.115001] [Medline: 31741352]
34. PortaMon. URL: https://www.artinis.com/portamon [accessed 2020-08-07]
35. What does the Everion measure?. URL: https://support.biofourmis.com/hc/en-us/articles/213613165-What-does-the-Everion-measure- [accessed 2020-08-07]
36. Abay TY, Kyriacou PA. Reflectance photoplethysmography as noninvasive monitoring of tissue blood perfusion. IEEE Trans Biomed Eng 2015 Sep;62(9):2187-2195. [doi: 10.1109/TBME.2015.2417863] [Medline: 25838515]
37. Lee H, Kim E, Lee Y, Kim H, Lee J, Kim M, et al. Toward all-day wearable health monitoring: an ultralow-power, reflective organic pulse oximetry sensing patch. Sci Adv 2018 Nov;4(11):eaas9530 [FREE Full text] [doi: 10.1126/sciadv.aas9530] [Medline: 30430132]
38. Validation of Software for Assessment of Sleep Apnea From Data Acquired by a Wearable Smartwatch. URL: https://clinicaltrials.gov/ct2/show/NCT03775291?term=smartwatch&cond=Sleep+Apnea&draw=2&rank=1 [accessed 2020-07-07]
39. Toronov VY, Zhang X, Webb AG. A spatial and temporal comparison of hemodynamic signals measured using optical and functional magnetic resonance imaging during activation in the human primary visual cortex. NeuroImage 2007 Feb;34(3):1136-1148. [doi: 10.1016/j.neuroimage.2006.08.048]
40. Toronov V, Webb A, Choi JH, Wolf M, Safonova L, Wolf U, et al. Study of local cerebral hemodynamics by frequency-domain near-infrared spectroscopy and correlation with simultaneously acquired functional magnetic resonance imaging. Opt Express 2001 Oct 08;9(8):417-427. [doi: 10.1364/oe.9.000417] [Medline: 19424359]
41. Gatto R, Hoffman W, Mueller M, Flores A, Valyi-Nagy T, Charbel FT. Frequency domain near-infrared spectroscopy technique in the assessment of brain oxygenation: a validation study in live subjects and cadavers. J Neurosci Methods 2006 Oct 30;157(2):274-277. [doi: 10.1016/j.jneumeth.2006.04.013] [Medline: 16730069]
42. Fantini S, Sassaroli A. Frequency-domain techniques for cerebral and functional near-infrared spectroscopy. Front Neurosci 2020;14:300 [FREE Full text] [doi: 10.3389/fnins.2020.00300] [Medline: 32317921]

XSL•FO
RenderX

43. Mensen A, Zhang Z, Qi M, Khatami R. The occurrence of individual slow waves in sleep is predicted by heart rate. Sci Rep 2016 Jul 22;6:29671 [FREE Full text] [doi: 10.1038/srep29671] [Medline: 27445083]
44. Berry R, Parish J, Hartse KM. The use of auto-titrating continuous positive airway pressure for treatment of adult obstructive sleep apnea. An American Academy of Sleep Medicine review. Sleep 2002 Mar 15;25(2):148-173. [Medline: 11902425]
45. Fantini S, Franceschini MA, Maier J, Walker S, Barbieri B, Gratton E. Frequency-domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry. Optical Engineering 1995 Jan 01;34(1):32. [doi: 10.1117/12.183988]
46. Villringer A, Chance B. Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci 1997 Oct;20(10):435-442. [doi: 10.1016/s0166-2236(97)01132-6] [Medline: 9347608]
47. Delpy DT, Cope M, van der Zee P, Arridge S, Wray S, Wyatt J. Estimation of optical pathlength through tissue from direct time of flight measurement. Phys Med Biol 1988 Dec;33(12):1433-1442. [doi: 10.1088/0031-9155/33/12/008] [Medline: 3237772]
48. Scholkmann F, Kleiser S, Metz AJ, Zimmermann R, Mata Pavia J, Wolf U, et al. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. NeuroImage 2014 Jan;85:6-27. [doi: 10.1016/j.neuroimage.2013.05.004]
49. Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Article in German. Annalen der Physik 1852;162(5):78-88. [doi: 10.1002/andp.18521620505]
50. Chan ED, Chan MM, Chan MM. Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations. Respiratory Medicine 2013 Jun;107(6):789-799. [doi: 10.1016/j.rmed.2013.02.004]
51. König V, Huch R, Huch A. Reflectance pulse oximetry--principles and obstetric application in the Zurich system. J Clin Monit Comput 1998 Aug;14(6):403-412. [doi: 10.1023/a:1009983010772] [Medline: 10023837]
52. Nitzan M, Romem A, Koppel R. Pulse oximetry: fundamentals and technology update. Med Devices (Auckl) 2014;7:231-239 [FREE Full text] [doi: 10.2147/MDER.S47319] [Medline: 25031547]
53. Scholkmann F, Wolf M. General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. J Biomed Opt 2013 Oct 01;18(10):105004. [doi: 10.1117/1.jbo.18.10.105004]
54. Franceschini MA, Gratton E, Fantini S. Noninvasive optical method of measuring tissue and arterial saturation: an application to absolute pulse oximetry of the brain. Opt Lett 1999 Jun 15;24(12):829-831. [doi: 10.1364/ol.24.000829] [Medline: 18073868]
55. Wolf M, Franceschini M, Paunescu L, Toronov V, Michalos A, Wolf U, et al. Absolute frequency-domain pulse oximetry of the brain: methodology and measurements. Adv Exp Med Biol 2003;530:61-73. [doi: 10.1007/978-1-4615-0075-9_7] [Medline: 14562705]
56. Vetter R, Rossini L, Ridolfi A, Sola J, Chetelat O, Correvon M, et al. Frequency Domain SpO2 Estimation Based on Multichannel Photoplethysmographic Measurements at the Sternum. Berlin, Heidelberg: Springer; 2009 Presented at: World Congress on Medical Physics and Biomedical Engineering; September 7 - 12, 2009; Munich. [doi: 10.1007/978-3-642-03882-2_86]
57. Zhao Y, Applegate MB, Istfan R, Pande A, Roblyer D. Quantitative real-time pulse oximetry with ultrafast frequency-domain diffuse optics and deep neural network processing. Biomed Opt Express 2018 Dec 01;9(12):5997-6008 [FREE Full text] [doi: 10.1364/BOE.9.005997] [Medline: 31065408]
58. Fantini S, Franceschini MA, Fishkin JB, Barbieri B, Gratton E. Quantitative determination of the absorption spectra of chromophores in strongly scattering media: a light-emitting-diode based technique. Appl Opt 1994 Aug 01;33(22):5204. [doi: 10.1364/ao.33.005204]
59. Fantini S, Hueber D, Franceschini MA, Gratton E, Rosenfeld W, Stubblefield PG, et al. Non-invasive optical monitoring of the newborn piglet brain using continuous-wave and frequency-domain spectroscopy. Phys Med Biol 1999 Jan 01;44(6):1543-1563. [doi: 10.1088/0031-9155/44/6/308]
60. Fantini S, Franceschini MA, Gratton E. Semi-infinite-geometry boundary problem for light migration in highly scattering media: a frequency-domain study in the diffusion approximation. J Opt Soc Am B 1994 Oct 01;11(10):2128. [doi: 10.1364/josab.11.002128]
61. Kleiser S, Nasseri N, Andresen B, Greisen G, Wolf M. Comparison of tissue oximeters on a liquid phantom with adjustable optical properties. Biomed Opt Express 2016 Aug 01;7(8):2973-2992 [FREE Full text] [doi: 10.1364/BOE.7.002973] [Medline: 27570691]
62. Stankovic M, Maulik D, Rosenfeld W, Stubblefield P, Kofinas A, Drexler S, et al. Real-time optical imaging of experimental brain ischemia and hemorrhage in neonatal piglets. J Perinat Med 1999;27(4):279-286. [doi: 10.1515/JPM.1999.039] [Medline: 10560079]
63. Fantini S, Franceschini MA, Gratton E, Hueber D, Rosenfeld W, Maulik D, et al. Non-invasive optical mapping of the piglet brain in real time. Opt Express 1999 Apr 12;4(8):308-314. [doi: 10.1364/oe.4.000308] [Medline: 19396287]
64. Hallacoglu B, Sassaroli A, Wysocki M, Guerrero-Berroa E, Schnaider Beeri M, Haroutunian V, et al. Absolute measurement of cerebral optical coefficients, hemoglobin concentration and oxygen saturation in old and young adults with near-infrared spectroscopy. J Biomed Opt 2012 Aug;17(8):081406-081401 [FREE Full text] [doi: 10.1117/1.JBO.17.8.081406] [Medline: 23224167]

XSL•FO
RenderX

65. Kleiser S, Ostojic D, Andresen B, Nasseri N, Isler H, Scholkmann F, et al. Comparison of tissue oximeters on a liquid phantom with adjustable optical properties: an extension. Biomed Opt Express 2018 Jan 01;9(1):86-101 [FREE Full text] [doi: 10.1364/BOE.9.000086] [Medline: 29359089]
66. SOMANETICS INVOS 3100A CEREBRAL OXIMETER (INVOS). U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K971628
67. FORE-SIGHT ABSOLUTE CEREBRAL AND SOMATIC OXIMETER, MC 2000 SERIES, MODELS MC2000, MC2010, MC2020, MC2030. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K094030
68. Franceschini MA, Thaker S, Themelis G, Krishnamoorthy KK, Bortfeld H, Diamond SG, et al. Assessment of infant brain development with frequency-domain near-infrared spectroscopy. Pediatr Res 2007 May;61(5, Part 1):546-551. [doi: 10.1203/pdr.0b013e318045be99]
69. Zhang Z, Bolz N, Laures M, Oremek M, Schmidt C, Qi M, et al. Cerebral blood volume and oxygen supply uniformly increase following various intrathoracic pressure strains. Sci Rep 2017 Aug 21;7(1):8344 [FREE Full text] [doi: 10.1038/s41598-017-08698-0] [Medline: 28827669]
70. Zhang Z, Schneider M, Laures M, Qi M, Khatami R. The comparisons of cerebral hemodynamics induced by obstructive sleep apnea with arousal and periodic limb movement with arousal: a pilot NIRS study. Front Neurosci 2016;10:403 [FREE Full text] [doi: 10.3389/fnins.2016.00403] [Medline: 27630539]
71. Zhang Z, Khatami R. Predominant endothelial vasomotor activity during human sleep: a near-infrared spectroscopy study. Eur J Neurosci 2014 Aug 26;40(9):3396-3404. [doi: 10.1111/ejn.12702]
72. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American Statistical Association 1988 Sep;83(403):596-610. [doi: 10.1080/01621459.1988.10478639]
73. Watzman H, Kurth C, Montenegro L, Rome J, Steven J, Nicolson SC. Arterial and venous contributions to near-infrared cerebral oximetry. Anesthesiology 2000 Oct;93(4):947-953 [FREE Full text] [doi: 10.1097/00000542-200010000-00012] [Medline: 11020744]
74. Abay TY, Kyriacou PA. Photoplethysmography for blood volumes and oxygenation changes during intermittent vascular occlusions. J Clin Monit Comput 2017 May 25;32(3):447-455. [doi: 10.1007/s10877-017-0030-2]
75. REGIONAL OXIMETER. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K102715
76. INSPECTRA TISSUE SPECTROMETER SYSTEM (INSPECTRA), MODEL 325. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K012759
77. VIOPTIX ODISSEY TISSUE OXIMETER TISSUE OXIMETER, MODEL OXY-2. U.S. Food and Drug Administration (FDA) 510(k) Database. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K042657
78. Benni PB, MacLeod D, Ikeda K, Lin H. A validation method for near-infrared spectroscopy based tissue oximeters for cerebral and somatic tissue oxygen saturation measurements. J Clin Monit Comput 2018 Apr;32(2):269-284 [FREE Full text] [doi: 10.1007/s10877-017-0015-1] [Medline: 28374103]
79. Kreeger RN, Ramamoorthy C, Nicolson SC, Ames WA, Hirsch R, Peng LF, et al. Evaluation of pediatric near-infrared cerebral oximeter for cardiac disease. Ann Thorac Surg 2012 Nov;94(5):1527-1533. [doi: 10.1016/j.athoracsur.2012.05.096] [Medline: 22858270]
80. Bickler PE, Feiner JR, Rollins MD. Factors affecting the performance of 5 cerebral oximeters during hypoxia in healthy volunteers. Anesth Analg 2013 Oct;117(4):813-823. [doi: 10.1213/ANE.0b013e318297d763] [Medline: 24023027]
81. Henson L, Calalang C, Temp J, Ward DS. Accuracy of a cerebral oximeter in healthy volunteers under conditions of isocapnic hypoxia. Anesthesiology 1998 Jan;88(1):58-65 [FREE Full text] [doi: 10.1097/00000542-199801000-00011] [Medline: 9447856]
82. Shah N, Trivedi NK, Clack SL, Shah M, Shah PP, Barker S. Impact of hypoxemia on the performance of cerebral oximeter in volunteer subjects. J Neurosurg Anesthesiol 2000 Jul;12(3):201-209. [doi: 10.1097/00008506-200007000-00001] [Medline: 10905567]
83. Benni P, Chen B, Dykes F, Wagoner S, Heard M, Tanner A, et al. Validation of the CAS neonatal NIRS system by monitoring vv-ECMO patients: preliminary results. Adv Exp Med Biol 2005;566:195-201. [doi: 10.1007/0-387-26206-7_27] [Medline: 16594153]
84. Rais-Bahrami K, Rivera O, Short BL. Validation of a noninvasive neonatal optical cerebral oximeter in veno-venous ECMO patients with a cephalad catheter. J Perinatol 2006 Aug 10;26(10):628-635. [doi: 10.1038/sj.jp.7211573]
85. Ikeda K, MacLeod DB, Grocott HP, Moretti EW, Ames W, Vacchiano C. The accuracy of a near-infrared spectroscopy cerebral oximetry device and its potential value for estimating jugular venous oxygen saturation. Anesth Analg 2014;119(6):1381-1392. [doi: 10.1213/ane.0000000000000463]
86. Redford D, Paidy S, Kashif F. Absolute and trend accuracy of a new regional oximeter in healthy volunteers during controlled hypoxia. Anesthesia & Analgesia 2014;119(6):1315-1319. [doi: 10.1213/ane.0000000000000474]
87. Franceschini MA, Boas DA, Zourabian A, Diamond SG, Nadgir S, Lin DW, et al. Near-infrared spiroximetry: noninvasive measurements of venous saturation in piglets and human subjects. J Appl Physiol (1985) 2002 Jan 01;92(1):372-384 [FREE Full text] [doi: 10.1152/jappl.2002.92.1.372] [Medline: 11744680]

XSL•FO
RenderX

88. MacDonald MJ, Tarnopolsky MA, Green HJ, Hughson RL. Comparison of femoral blood gases and muscle near-infrared spectroscopy at exercise onset in humans. J Appl Physiol (1985) 1999 Feb;86(2):687-693 [FREE Full text] [doi: 10.1152/jappl.1999.86.2.687] [Medline: 9931209]
89. Lee H, Ko H, Lee J. Reflectance pulse oximetry: practical issues and limitations. ICT Express 2016 Dec;2(4):195-198. [doi: 10.1016/j.icte.2016.10.004]
90. Tamura T. Current progress of photoplethysmography and SPO for health monitoring. Biomed Eng Lett 2019 Feb;9(1):21-36 [FREE Full text] [doi: 10.1007/s13534-019-00097-w] [Medline: 30956878]
91. Mendelson Y, Ochs B. Noninvasive pulse oximetry utilizing skin reflectance photoplethysmography. IEEE Trans Biomed Eng 1988 Oct;35(10):798-805. [doi: 10.1109/10.7286] [Medline: 3192229]
92. Hickey M, Kyriacou PA. Optimal spacing between transmitting and receiving optical fibres in reflectance pulse oximetry. In: J Phys: Conf Ser. 2007 Nov 07 Presented at: Third International Conference on Optical and Laser Diagnostics (ICOLAD 2007); 22–25 May 2007; City University London p. 012030. [doi: 10.1088/1742-6596/85/1/012030]
93. Pittman S, Ayas N, MacDonald M, Malhotra A, Fogel R, White P. Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation. Sleep 2004 Aug 01;27(5):923-933 [FREE Full text] [doi: 10.1093/sleep/27.5.923] [Medline: 15453551]

Abbreviations

AHI: Apnea–Hypopnea Index
arm_OD: arm oxygen desaturation
AVR: arterio-venous ratio
CPAP: continuous positive airway pressure
CW-NIRS: continuous wave near-infrared spectroscopy
DPF: differential path-length factor
FDMD: frequency-domain multidistance
finger_OD: fingertip oxygen desaturation
HbO_2: oxygenated hemoglobin
HHb: deoxygenated hemoglobin
LoA: limit of agreement
MBLL: modified Beer–Lambert law
NIRS: near-infrared spectroscopy
OD: oxygen desaturation
OSA: obstructive sleep apnea
PSG: polysomnography
R-PPG: reflectance photoplethysmography
SaO_2: arterial oxygen saturation
SNR: signal-to-noise ratio
SpO_2: peripheral capillary oxygen saturation
StO_2: peripheral tissue oxygen saturation
SvO_2: venous oxygen saturation
T-PPG: transmission photoplethysmography

Edited by R Kukafka; submitted 07.09.20; peer-reviewed by A Vehkaoja, T Penzel; comments to author 20.10.20; revised version received 26.02.21; accepted 24.05.21; published 29.07.21.

Please cite as:
Zhang Z, Qi M, Hügli G, Khatami R
The Challenges and Pitfalls of Detecting Sleep Hypopnea Using a Wearable Optical Sensor: Comparative Study
J Med Internet Res 2021;23(7):e24171
URL: https://www.jmir.org/2021/7/e24171
doi:10.2196/24171
PMID:34326039

©Zhongxing Zhang, Ming Qi, Gordana Hügli, Ramin Khatami. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any

XSL•FO
RenderX

medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Empowering Patients Living With Chronic Conditions Using Video as an Educational Tool: Scoping Review

Olga Navarro[1,2]; Marta Escrivá[1]; Raquel Faubel[3,4,5], PhD; Vicente Traver[2,4], PhD

[1]Department of Nursing, Catholic University of Valencia, Valencia, Spain

[2]Institute of Information and Communication Technologies, Universitat Politècnica de València, Valencia, Spain

[3]Department of Physiotherapy, Universitat de València, Valencia, Spain

[4]Unidad Mixta de Reingeniería de Procesos Sociosanitarios, IIS La Fe-Universitat Politècnica de València, Valencia, Spain

[5]Physiotherapy in Motion, Multispeciality Research Group, Department of Physiotherapy, Universitat de València, Valencia, Spain

Corresponding Author:
Raquel Faubel, PhD
Department of Physiotherapy
Universitat de València
Gascó Oliag 5
Valencia, 46010
Spain
Phone: 34 963983853
Email: raquel.faubel@uv.es

Abstract

Background: Video is used daily for various purposes, such as leisure, culture, and even learning. Currently, video is a tool that is available to a large part of the population and is simple to use. This audio-visual format has many advantages such as its low cost, speed of dissemination, and possible interaction between users. For these reasons, it is a tool with high dissemination and educational potential, which could be used in the field of health for learning about and management of chronic diseases by adult patients.

Objective: The following review determines whether the use of health educational videos by adult patients with chronic diseases is effective for their self-management according to the literature.

Methods: An electronic literature search of the PubMed, CINAHL, and MEDLINE (via the EBSCOhost platform) databases up to April 2020 was conducted. The systematic scoping review followed the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) methodology.

Results: After reviewing 1427 articles, 12 were selected as the most consistent with the proposed inclusion criteria. After their review, it was found that the studies showed that video is effective as a tool for improving care related to chronic diseases.

Conclusions: Video is effective in improving the care and quality of life for patients with chronic diseases, whether the initiative for using video came from their health care professionals or themselves.

(J Med Internet Res 2021;23(7):e26427) doi:10.2196/26427

KEYWORDS

patients; health education; self-care; video; chronic disease

Introduction

Overview

The increase in the use of social networks and the need for patients to know more about their disease or that of a loved one and how to manage it properly have led to calls for the health system to be updated and health professionals to offer patients reliable and quality tools. Video has become a powerful teaching and dissemination method in all fields. In the health field, it is also used for educational and empowering purposes for the patient, especially when the patient has a chronic disease.

By supplying educational videos for viewing in the health field, the aim is to help patients improve their quality of life through more rigorous self-care, encouraging their access to the health system and often, making personalized care available if video enables live video conferences with health care professionals. Video is a very useful and effective tool, both for professionals

XSL•FO
RenderX

and patients, due to its low cost, speed of dissemination, and ease of access.

Background

The annual report "The Global State of Digital in 2019" created by Hootsuite and We Are Social [1] positioned the YouTube audio-visual platform as the most used social network in Spanish-speaking countries (Mexico, Colombia, Argentina, and Spain). In Spain, 89% of social network users use YouTube [2]. One of the characteristics of the videosocial networks (eg, YouTube or Vimeo) is that the shared videos can host comments and these are registered publicly; in this way, a question that a person asks publicly can be read by other users who have a similar question. This feature, together with its enormous popularity, makes it a tool worth considering for disseminating health information [3,4].

Numerous studies consider that online video content on health is useful and highly effective for educating patients [5,6], although others point out that this content should be viewed with caution since it could be erroneous or confusing and not provide quality information [7-10]. In fact, the most popular videos or those with the highest number of views may contain low-quality or even inappropriate content.

For this reason, many researchers agree on the need for professionals to lead the creation of quality video content [4], although it could also be helpful for patients to do so [11].

It is important to bear in mind that YouTube or Vimeo have their own algorithm for classifying videos that are published, so that, despite efforts from professionals or institutions to generate quality videos, such videos may not reach the target public if they are not properly disseminated and other approaches are not used [12]. Given the lack of quality or patient-specific content on generalist platforms such as YouTube and Vimeo, the main role of the health professional should be that of content curator or link supplier, helping patients to select quality resources [13,14].

The objective of this scoping review is to determine whether the use of educational videos in the health field for self-management of chronic diseases by adult patients is effective.

Methods

This study follows the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) [15]. The PICO (Patient, Intervention, Comparison, Outcomes) framework was used to answer the question: "Does access to audio-visual tools improve the care of patients and/or their families as a way of empowering them to face up to their pathologies?" as shown in Table 1.

Table 1. The PICO (Patient, Intervention, Comparison, Outcomes) framework.

PICO framework	Description	Application to this study
P	Definition of the problem or patient	Adult population with chronic diseases or their adult caregivers
I	Interventions	Using video to obtain information about their own illness or a family member's illness
C	Comparison	The effectiveness of viewing videos to improve the care of the chronic pathology to be treated
O	Outcomes= Results	Management of chronic disease in adult patient after the use of video as an educational tool.

The search for journal articles used for this bibliographic review was carried out in the following databases: CINAHL and MEDLINE (via the EBSCOhost platform) and PubMed. The search terms proposed in the process of searching and selecting articles in both databases and the results obtained, in the form of articles, respectively, are shown in the following paragraphs.

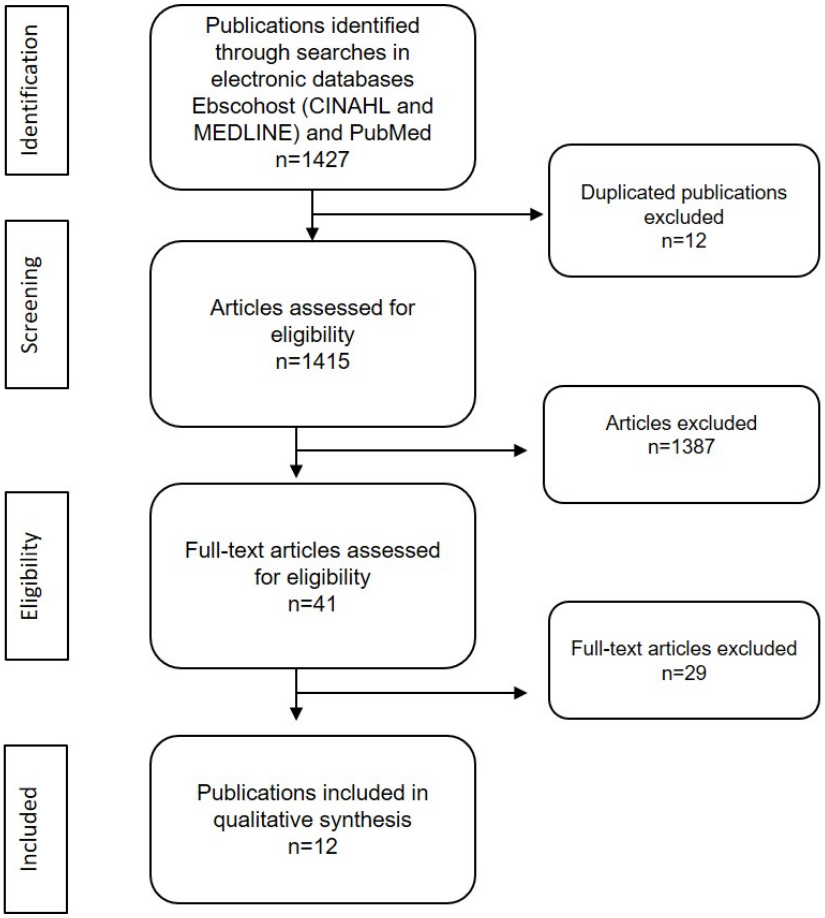
The following keywords (MESH descriptors) were used to search for articles: patients, adult, young adult, family, education, health education, self-care, power (psychological), audio-visual aids, video recording, webcasts, chronic disease, chronic pain, caregivers, nursing models. The keywords in Spanish (DECS descriptors) were the following: Pacientes, paciente crónico, familia, cuidadores, educación, autocuidado, vídeo, enfermería, audiovisual, tratamiento, empoderamiento.

To perform this bibliographic review, we selected articles using the following inclusion criteria: articles published between the years 2017 and 2020 (both included); articles dealing with chronic diseases, adult patients, or caregivers; articles where audio-visual content is the main educational tool; studies carried out in humans; text in English and Spanish. All types of sources (academic publications, book publications, reports, and dissertations) included in these databases were accepted. Concerning exclusion criteria, articles on acute diseases or those conducted with pediatric patients were excluded.

As shown in the PRISMA flow diagram (Figure 1), after the initial search and eliminating duplicates, 1415 articles were identified, of which 1387 were eliminated after reading the title and summary. Of the 41 remaining, after critical reading of the complete text, 29 other articles were rejected, and 12 studies were finally selected for inclusion in the scoping review.

XSL•FO
RenderX

Figure 1. Results of the study selection process using the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) methodology.



Methodological quality was assessed using the Downs and Black Checklist for Quality Assessment [16] in the modified version used in several previous studies [17-19]. This checklist is made up of a total of 27 questions classified in 4 domains (reporting, external validity, internal validity, and power). Scores on the modified Downs and Black checklist were classified as "excellent" (score 24-28), "good" (19-23), "fair" (14-18), or "poor" (≤13).

Relevant data regarding study design, characteristics of participants including chronic disease, and interventions carried out including characteristics of the video used as an educational tool were extracted. Specifically, regarding video characteristics, information was collected about the source of the video (created ad hoc for the study or already existing videos) and how it was used as an educational tool (self-administered, shown during an outpatient visit, or shown during a video conference in a synchronous form). Information about results was also extracted related to the expected results previously described in each study. Positive results were considered when a significant improvement in the objectives proposed for the study for the educational tool was shown.

Results

In the 12 studies selected to carried out this study, a total of 1398 participants were included. The sample size of the different studies ranged from 8 participants [20] to 429 participants [21], with an average size of 116.5 participants per study. Of the 12 studies, 8 studies included more than 50 subjects. Table 2 summarizes the most relevant data from the articles included.

XSL•FO
RenderX

Table 2. Descriptive characteristics of the included studies.

Author(s), year	Title or objective	Study design	Participants	Intervention	Results	Downs and Black Checklist score
Albert et al, 2017 [22]	Factors associated with telemonitoring use among patients with chronic heart failure	Descriptive cross-sectional study	206 outpatients and hospitalized patients with heart failure	A 6-minute video was offered that dealt with telemonitoring, which was intended to collect information on the patient's blood pressure and weight to send to the caregiver. Subsequently, questionnaires were used to evaluate the efficacy.	The intervention was well received by the patients, especially when satisfying the personal needs of care and learning in a bidirectional way with a doctor or a nurse, through the use of smartphones.	16
Rosen et al, 2017 [23]	Telehealth protocol to prevent readmission among high-risk patients with congestive heart failure	Quasiexperimental study	50 patients with congestive heart failure	A telehealth platform was developed that enables, through videoconferencing, educating patients about their disease, to prevent hospital readmissions for congestive heart failure.	Adherence to treatment was increased, and there was a marked decrease in readmission of patients with congestive heart failure.	19
Vogler et al, 2017 [24]	Assessing outcomes of educational videos in group visits for patients with chronic pain at an academic primary care clinic	Analytical observational prospective cohort study	14 patients with chronic noncancerous pain who underwent an educational program on pain	Educational videos were offered to patients and discussed orally in a subsequent group visit. In total, there were 4 group visits.	The study participants, by improving their knowledge, reduced the doses of painkillers they took, in addition to reducing visits to the emergency room for pain.	17
Farver-Vestergaard et al, 2019 [20]	Teledelivered mindfulness-based cognitive therapy in chronic obstructive pulmonary disease: a mixed methods feasibility study	Quasiexperimental study	8 patients with COPD[a] underwent Mindfulness-Based Cognitive Therapy through videoconferencing	Two groups underwent mindfulness therapy to reduce psychological distress and improve the physical health of COPD through video conference sessions.	Clinical improvement was observed in hospital depression and anxiety. The patients perceived an improvement in unpleasant physical sensations and psychological symptoms.	18
Ward et al, 2018 [25]	Evaluation of multidisciplinary pulmonary rehabilitation education delivered by either DVD or spoken talk	Analytical observational cohort study	123 patients with COPD in a pulmonary rehabilitation study	Two groups were divided: One was provided with education about pulmonary rehabilitation through DVD, and the other group received the same information through oral and face-to-face discussion.	Education via DVD was found to be as effective as traditional education.	16
Ketelaars et al, 2017 [26]	The effect of video information on anxiety levels in women attending colposcopy: a randomized controlled trial	Randomized controlled clinical trial	136 women, older than 18 years, with a positive hrHPV[b] test, referred for colposcopy	Group A received information about colposcopy through a video and a brochure. Group B was only given the brochure.	The video did not significantly reduce the levels of anxiety, depression, or pain in the participants. But participants positively valued the video information.	17
De Lepeleere et al, 2017 [27]	The effect of an online video intervention “movie models” on specific parenting practices and parental self-efficacy related to children’s physical activity, screen-time and healthy diet	Quasiexperimental two-arm study	238 parents with children 6-12 years old	The study offered 22 online, 2-minute videos on obesity and chronicity prevention for 4 weeks.	Parents valued the video as a useful and applicable tool. It was an effective tool for improving family habits and parental self-efficacy.	20

XSL•FO
RenderX

Author(s), year	Title or objective	Study design	Participants	Intervention	Results	Downs and Black Checklist score
Bakas et al, 2019 [28]	Using telehealth to optimize healthy independent living for older adults: a feasibility study	Quasiexperimental study	22 older adults with some chronic health condition	A textbook, advice sheets, and 2 DVDs were provided; 3 telepresence sessions were held where the patients were trained using the tools provided.	Improvements were found in quality of life, self-efficacy, and confidence perceived by the patients.	17
Zanaboni et al, 2017 [29]	Long-term exercise maintenance in COPD via telerehabilitation: a 2-year pilot study	Quasiexperimental study	10 adult COPD patients	The intervention consisted of providing patients with exercises at home, supervised by videoconference by a physiotherapist.	It was determined that telerehabilitation is feasible for maintaining good long-term health status in COPD patients.	20
Taylor et al, 2018 [30]	Integrating innovative telehealth solutions into an interprofessional team-delivered chronic care management pilot program	Retrospective observational study	69 patients with 3 or more chronic pathologies, taking at least 5 drugs	A teleconsultation service was provided with a pharmacist to review treatments, doses, and improve adherence.	One-third of the patients changed their habits after pharmaceutical advice. It was found to be a useful tool for reducing errors.	15
McLeod et al, 2020 [21]	Impact of a comprehensive digital health programme on HbA1c and weight after 12 months for people with diabetes and prediabetes: a randomised controlled trial	Randomized controlled trial	429 patients with diabetes not taking insulin and daily access to the internet	The control and intervention arms received usual care. The intervention arm received the BetaMe/Melon program over 12 months, delivered through mobile devices and web-based platforms.	There were small improvements in HbA1c[c] and weight at 4 months that had largely attenuated by 12 months. The BetaMe/ Melon program in its current form cannot be recommended for use in the management of diabetes or prediabetes.	25
Locke et al, 2019 [31]	Using video telehealth to facilitate inhaler training in rural patients with obstructive lung disease	Retrospective observational analytical cohort study	93 resident patients with COPD or asthma in a rural setting	Live video training sessions were given to patients to explain the use of inhalers.	Improvement in inhalation technique was achieved by patients with asthma or COPD.	19

[a]COPD: chronic obstructive pulmonary disease.

[b]hrHPV: high-risk human papillomavirus.

[c]HbA1c: glycosylated hemoglobin.

All videos employed as educational tools were specifically created within the frame of research. None of the studies included in the review used public videos or videos already hosted on a YouTube channel.

Regarding the study design, 5 of the studies were designed as a quasiexperimental, pre-post study. Another 5 studies were observational with either a retrospective, prospective, or descriptive design. In those studies, the objective was to assess the impact of videos as an educational tool provided to a group of patients for training or follow-up. Only 2 of the studies were designed as a randomized controlled trial.

Regarding the chronic disease of the participants, the selected studies included participants with diagnoses of chronic obstructive pulmonary disease (COPD; 4/12, 33%) [20,25,29,31], which was the most frequent; heart failure (2/12, 17%) [22,23]; chronic pain (1/12, 8%) [24]; diabetes (1/12, 8%) [21]; or squamous intraepithelial lesion due to human papillomavirus (HPV; 1/12, 8%) [26]. Participants with any chronic conditions were included in 3 of the studies [27,28,30], while quality of life was measured in 25% (3/12) of the studies.

Each study included in the review provided different educational video tools to the subjects who participated in the research: educational video, informative video, video based on real cases, explanatory video. They were provided in different ways and in some cases combined with other educational strategies. Specifically, some studies used videoconferencing to provide physical therapy exercises [29], mindfulness therapy [20], pharmacological advice [30], or training about using inhalers [31]. Other studies used online videos [26,27] for different purposes such as education for pain management [24], information about diagnostic procedures [26], or disease prevention [27]. One study took advantage of an outpatient visit or hospitalization for the patient to view 1 short video about telemonitoring options for their disease [22]. In other studies, videos were provided in a DVD delivered to the patient independently [25] or combined with other activities such as

XSL•FO
RenderX

videoconference sessions [28]. Lastly, some studies included an educational video combined with other tools incorporated in a complex platform for telehealth [21,23].

These studies assessed whether the application of video had improved the management of chronic disease, and successful results were obtained in most cases. Of the 12 studies included in the review, 2 studies [21,26] did not show significant results to support the hypothesis that video is an effective tool to improve the health of patients with chronic disease in the long term. The study carried out by Ketelaars et al [26] found that the educational video did not significantly reduce the anxiety levels of the participants; that is, it was not effective as a stress reduction tool for women with HPV before a diagnostic procedure, compared with a brochure. However, the patients positively valued the videos offered. On the other hand, McLeod et al [21] found positive results on different outcomes in the first few months of a complex intervention for patients with diabetes. Nevertheless, after 12 months of follow-up, there were no significant differences.

Positive significant results were found in 10 of the 12 studies with a great heterogeneity of outcome variables: The doses of drugs and analgesics [24,30], exacerbations of chronic diseases, readmissions, and emergency room visits [23,24] were all reduced. Studies aimed specifically at patients with COPD showed that rehabilitation using video resources is also effective [25,29,31]. In addition to this, studies also found a decrease in pain, anxiety, and depression [20].

Lastly, improvements in disease knowledge [25], satisfaction [27], and health-related quality of life [28] were also found. These data indicate the effectiveness of the audio-visual tools provided to patients for improving management of chronic diseases and reinforcing health advice.

In terms of methodological quality, according to the modified version of the Downs and Black Checklist for Quality Assessment (Multimedia Appendix 1), 7 of the studies presented with a fair quality grade (scores from 15 to 18), and the other 5 studies had excellent or good methodological quality (scores from 19 to 25). The average score was 18.25. Items with the worst ratings on the quality scale were those related to blinding and randomization.

Discussion

Principal Findings

This scoping review identified 1427 articles, of which 12 fulfilled the selection criteria to determine the effectiveness of video as an educational and empowering tool for patients with chronic disease and their caregivers. The results of the review showed that audio-visual instruments are highly effective in the acquisition of competencies by the patient for self-management of chronic diseases. According to the studies reviewed, patients and caregivers themselves were able to improve their self-care and management of chronic diseases after complementing health advice with specific video resources.

In all the articles analyzed in this review, professionals offered the patient or group of patients videos or different audio-visual instruments specifically developed for research. Data on videos freely consulted on the internet by the patient were not included in these articles, nor were different video formats (eg, videoconferencing, DVD viewing, informative or explanatory video, educational video) and their effectiveness compared.

The use of educational videos could contribute to improving the general health status of patients with chronic disease and could even act at a preventive level, reducing the number of admissions and hospital stays. Along these lines, the use of educational videos with adult patients with chronic disease could help reduce anxiety, depression, pain, and even rescue medication doses that patients routinely take.

In addition, educational videos could be a powerful tool for the empowerment of patients with chronic diseases, helping to resolve their issues in a fast, economical, and dynamic way.

It was observed that in all the studies reviewed, health professionals directly offered the video or other audio-visual instruments to a group of patients, and patients did not freely search for videos on the internet. This issue omits investigations where a significant proportion of the population consults and shares videos on the internet, and thus, we were unable to affirm whether in this freer area, not guided by a professional, there are differences in effectiveness.

If audio-visual tools were used in the health system, patients could improve their health status and decrease their admissions and stays in hospital. In addition to this, they could resolve issues wherever they may be, in a fast, economical, and dynamic way.

The implementation of audio-visual tools could reduce the workload of health professionals, reach a greater number of people, decrease health spending, and carry out more effective health education, among other possible benefits.

In this review, in most of the studies, the professionals reinforced the videos with other types of supplementary educational material, such as infographics, brochures, and face-to-face visits.

Despite these results, video is not yet implemented in clinical practice as an educational tool of great value, perhaps due to the scarce evidence in this regard and the lack of training for professionals, among other factors.

Health professionals could have a relevant part in developing or selecting useful and high-quality video material for patients, particularly since videos available via services such as YouTube or Vimeo should be screened by an expert before patients view them, to increase the level of safety and trust in the content [32]. Indeed, material obtained in these video platforms without supervision could even contain dangerous information for patients' health [33,34].

For this reason, it would be useful for health professionals to be updated in the use of new technologies and more specifically video, since audio-visual instruments are highly effective in the management of chronic diseases. Despite this, these tools, in some of the studies that were reviewed, had to be reinforced with additional information that was intelligible by the entire population, including the elderly population.

XSL•FO
RenderX

Although audio-visual tools are effective for improving the management of pathologies, personalized attention should not be omitted when necessary, so that the combined effectiveness of all educational instruments persists over time and is retained in the population. Good communication between patients and health professionals, either face-to-face or using videoconference, will reduce the need to get additional information. That would avoid browsing video material with low scientific evidence that could be harmful for the patient [35].

The present scoping review was conducted following the PRISMA checklist. Despite this, one of the limitations we found was the low number of articles included in the review and the great heterogeneity in the design of the included studies and chronic diseases included. On the other hand, the methodological quality of some of the studies could be explained by the challenge of recruiting participants for educational research studies and designing randomized controlled trials with a long-term follow-up. At the same time, the diverse range of aspects related to chronicity analyzed in the included studies also represents that using video as an educational and empowering tool for patients with chronic disease could improve several aspects of the disease.

With regard to chronic diseases, after reviewing the 12 selected articles, we concluded that the authors attached great importance to a group of chronic diseases such as COPD, heart disease, and diabetes. On the other hand, there are various chronic diseases that have a high impact on the population and were not included in the studies (eg, hypertension, obesity, fibromyalgia, and dementias); therefore, it would be useful to include them in this type of research. Further possible benefits in the use of video might be found.

On the other hand, the authors suggested tools such as videoconferencing with health care professionals (ie, nurses, doctors, pharmacists) and showed that after use of such tools, significant improvements in patient education are obtained. However, these types of resources are not widely used in health services either. It would be appropriate to study the barriers that hinder the use of these tools when receiving and offering health information through videos, such as lack of expertise in use, lack of technical resources, security, or support problems from the health system.

Conclusions

Despite the conclusion of the effectiveness of video for educating patients and improving self-care for chronic diseases in different articles published in well-known and high-impact medical journals, the use of video in the health field has not yet been implemented on a routine basis. It would be advisable to continue researching in this area and identify the advantages and benefits that audio-visual instruments can bring to patients with chronic diseases, their caregivers, and health professionals themselves, consolidating video as a complementary tool and of great support in reinforcing the health advice offered in consultations and clinical events.

Acknowledgments

This study was supported in part by Universidad Católica San Vicente Mártir.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological quality according to the modified version of Downs and Black Checklist for Quality Assessment.
[DOCX File , 20 KB - jmir_v23i7e26427_app1.docx]

References

1. The global state of digital in 2019. Hootsuite. URL: https://www.hootsuite.com/resources/digital-in-2019 [accessed 2021-06-22]
2. Newberry C, Adame A. 22 Estadísticas de YouTube esenciales para este año. Hootsuite. 2019 May 15. URL: https://blog.hootsuite.com/es/estadisticas-de-youtube/ [accessed 2021-06-22]
3. Khatri P, Singh SR, Belani NK, Yeong YL, Lohan R, Lim YW, et al. YouTube as source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content. Travel Med Infect Dis 2020;35:101636 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101636] [Medline: 32205267]
4. Rubel KE, Alwani MM, Nwosu OI, Bandali EH, Shipchandler TZ, Illing EA, et al. Understandability and actionability of audiovisual patient education materials on sinusitis. Int Forum Allergy Rhinol 2020 Apr 12;10(4):564-571. [doi: 10.1002/alr.22518] [Medline: 31930714]
5. Altan Şallı G, Egil E. Are YouTube videos useful as a source of information for oral care of leukemia patients? Quintessence Int 2020;51(1):78-85. [doi: 10.3290/j.qi.a43665] [Medline: 31781693]
6. Ruiz-Roca JA, Martínez-Izquierdo A, Mengual-Pujante D, López EPF, López-Jornet P. Is YouTube a useful tool for oral care in patients with Parkinson's disease? Spec Care Dentist 2020 Sep 06;40(5):464-469. [doi: 10.1111/scd.12489] [Medline: 32628790]

XSL•FO
RenderX

7. Jain N, Abboudi H, Kalic A, Gill F, Al-Hasani H. YouTube as a source of patient information for transrectal ultrasound-guided biopsy of the prostate. Clin Radiol 2019 Jan;74(1):79.e11-79.e14. [doi: 10.1016/j.crad.2018.09.004] [Medline: 30322707]
8. Cassidy JT, Fitzgerald E, Cassidy ES, Cleary M, Byrne DP, Devitt BM, et al. YouTube provides poor information regarding anterior cruciate ligament injury and reconstruction. Knee Surg Sports Traumatol Arthrosc 2018 Mar 17;26(3):840-845. [doi: 10.1007/s00167-017-4514-x] [Medline: 28314888]
9. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube Useful as a Source of Health Information for Adults With Type 2 Diabetes? A South Asian Perspective. Can J Diabetes 2018 Aug;42(4):395-403.e4. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
10. Lambert K, Mullan J, Mansfield K, Koukomous A, Mesiti L. Evaluation of the quality and health literacy demand of online renal diet information. J Hum Nutr Diet 2017 Oct;30(5):634-645. [doi: 10.1111/jhn.12466] [Medline: 28211108]
11. Salama A, Panoch J, Bandali E, Carroll A, Wiehe S, Downs S, et al. Consulting "Dr. YouTube": an objective evaluation of hypospadias videos on a popular video-sharing website. J Pediatr Urol 2020 Feb;16(1):70.e1-70.e9 [FREE Full text] [doi: 10.1016/j.jpurol.2019.11.011] [Medline: 31928900]
12. Fernandez-Llatas C, Traver V, Borras-Morell J, Martinez-Millana A, Karlsen R. Are Health Videos from Hospitals, Health Organizations, and Active Users Available to Health Consumers? An Analysis of Diabetes Health Video Ranking in YouTube. Comput Math Methods Med 2017;2017:8194940-8194949 [FREE Full text] [doi: 10.1155/2017/8194940] [Medline: 28243314]
13. Al-Busaidi IS, Anderson TJ, Alamri Y. Qualitative analysis of Parkinson's disease information on social media: the case of YouTube™. EPMA J 2017 Sep;8(3):273-277 [FREE Full text] [doi: 10.1007/s13167-017-0113-7] [Medline: 29021838]
14. Lee JL, Frey M, Frey P, Hollin IL, Wu AW. Seeing is Engaging: Vlogs as a Tool for Patient Engagement. Patient 2017 Jun 18;10(3):267-270 [FREE Full text] [doi: 10.1007/s40271-017-0215-2] [Medline: 28101817]
15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
16. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
17. Trac MH, McArthur E, Jandoc R, Dixon SN, Nash DM, Hackam DG, et al. Macrolide antibiotics and the risk of ventricular arrhythmia in older adults. CMAJ 2016 Apr 19;188(7):E120-E129 [FREE Full text] [doi: 10.1503/cmaj.150901] [Medline: 26903359]
18. Hooper P, Jutai JW, Strong G, Russell-Minda E. Age-related macular degeneration and low-vision rehabilitation: a systematic review. Can J Ophthalmol 2008 Apr;43(2):180-187. [doi: 10.3129/i08-001] [Medline: 18347620]
19. O'Connor SR, Tully MA, Ryan B, Bradley JM, Baxter GD, McDonough SM. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. BMC Res Notes 2015 Jun 06;8:224 [FREE Full text] [doi: 10.1186/s13104-015-1181-1] [Medline: 26048813]
20. Farver-Vestergaard I, O'Connor M, Smith NC, Løkke A, Bendstrup E, Zachariae R. Tele-delivered mindfulness-based cognitive therapy in chronic obstructive pulmonary disease: A mixed-methods feasibility study. J Telemed Telecare 2018 Jun 25;25(8):468-475. [doi: 10.1177/1357633x18780563]
21. McLeod M, Stanley J, Signal V, Stairmand J, Thompson D, Henderson K, et al. Impact of a comprehensive digital health programme on HbA and weight after 12 months for people with diabetes and prediabetes: a randomised controlled trial. Diabetologia 2020 Dec;63(12):2559-2570. [doi: 10.1007/s00125-020-05261-x] [Medline: 32886192]
22. Albert NM, Dinesen B, Spindler H, Southard J, Bena JF, Catz S, et al. Factors associated with telemonitoring use among patients with chronic heart failure. J Telemed Telecare 2017 Feb;23(2):283-291. [doi: 10.1177/1357633X16630444] [Medline: 26869144]
23. Rosen D, McCall JD, Primack BA. Telehealth Protocol to Prevent Readmission Among High-Risk Patients With Congestive Heart Failure. Am J Med 2017 Nov;130(11):1326-1330. [doi: 10.1016/j.amjmed.2017.07.007] [Medline: 28756266]
24. Vogler CN, Sattovia S, Salazar LY, Leung TI, Botchway A. Assessing outcomes of educational videos in group visits for patients with chronic pain at an academic primary care clinic. Postgrad Med 2017 Jun 16;129(5):524-530. [doi: 10.1080/00325481.2017.1324228] [Medline: 28447542]
25. Ward S, Sewell L, Singh S. Evaluation of multidisciplinary pulmonary rehabilitation education delivered by either DVD or spoken talk. Clin Respir J 2018 Nov 25;12(11):2546-2550. [doi: 10.1111/crj.12954] [Medline: 30221825]
26. Ketelaars PJW, Buskes MHM, Bosgraaf RP, van Hamont D, Prins JB, Massuger LFAG, et al. The effect of video information on anxiety levels in women attending colposcopy: a randomized controlled trial. Acta Oncol 2017 Dec;56(12):1728-1733. [doi: 10.1080/0284186X.2017.1355108] [Medline: 28760058]
27. De Lepeleere S, De Bourdeaudhuij I, Cardon G, Verloigne M. The effect of an online video intervention 'Movie Models' on specific parenting practices and parental self-efficacy related to children's physical activity, screen-time and healthy diet: a quasi experimental study. BMC Public Health 2017 Apr 27;17(1):366 [FREE Full text] [doi: 10.1186/s12889-017-4264-1] [Medline: 28449658]

XSL•FO
RenderX

28. Bakas T, Sampsel D, Israel J, Chamnikar A, Bodnarik B, Clark JG, et al. Using telehealth to optimize healthy independent living for older adults: A feasibility study. Geriatr Nurs 2018 Sep;39(5):566-573 [FREE Full text] [doi: 10.1016/j.gerinurse.2018.04.002] [Medline: 29804689]
29. Zanaboni P, Hoaas H, Aarøen Lien L, Hjalmarsen A, Wootton R. Long-term exercise maintenance in COPD via telerehabilitation: a two-year pilot study. J Telemed Telecare 2016 Jul 09;23(1):74-82. [doi: 10.1177/1357633x15625545]
30. Taylor AM, Bingham J, Schussel K, Axon DR, Dickman DJ, Boesen K, et al. Integrating Innovative Telehealth Solutions into an Interprofessional Team-Delivered Chronic Care Management Pilot Program. J Manag Care Spec Pharm 2018 Aug;24(8):813-818. [doi: 10.18553/jmcp.2018.24.8.813] [Medline: 30058982]
31. Locke ER, Thomas RM, Woo DM, Nguyen EHK, Tamanaha BK, Press VG, et al. Using Video Telehealth to Facilitate Inhaler Training in Rural Patients with Obstructive Lung Disease. Telemed J E Health 2019 Mar;25(3):230-236 [FREE Full text] [doi: 10.1089/tmj.2017.0330] [Medline: 30016216]
32. Gabarron E, Fernandez-Luque L, Armayones M, Lau AY. Identifying Measures Used for Assessing Quality of YouTube Videos with Patient Health Information: A Review of Current Literature. Interact J Med Res 2013 Feb 28;2(1):e6 [FREE Full text] [doi: 10.2196/ijmr.2465] [Medline: 23612432]
33. Mueller SM, Hongler VNS, Jungo P, Cajacob L, Schwegler S, Steveling EH, et al. Fiction, Falsehoods, and Few Facts: Cross-Sectional Study on the Content-Related Quality of Atopic Eczema-Related Videos on YouTube. J Med Internet Res 2020 Apr 24;22(4):e15599 [FREE Full text] [doi: 10.2196/15599] [Medline: 32329744]
34. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
35. Langford A, Loeb S. Perceived Patient-Provider Communication Quality and Sociodemographic Factors Associated With Watching Health-Related Videos on YouTube: A Cross-Sectional Analysis. J Med Internet Res 2019 May 17;21(5):e13512 [FREE Full text] [doi: 10.2196/13512] [Medline: 31102372]

Abbreviations

COPD: Chronic obstructive pulmonary disease
HPV: human papillomavirus
PICO: Patient, Intervention, Comparison, Outcomes
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analysis

Edited by R Kukafka; submitted 11.12.20; peer-reviewed by A Langford, V Press; comments to author 06.02.21; revised version received 01.03.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Navarro O, Escrivá M, Faubel R, Traver V
Empowering Patients Living With Chronic Conditions Using Video as an Educational Tool: Scoping Review
J Med Internet Res 2021;23(7):e26427
URL: https://www.jmir.org/2021/7/e26427
doi:10.2196/26427
PMID:34255671

©Olga Navarro, Marta Escrivá, Raquel Faubel, Vicente Traver. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Pharmacy Customers' Experiences of Use, Usability, and Satisfaction of a Nationwide Patient Portal: Survey Study

Maria Sääskilahti[1], MSc; Riitta Ahonen[1], PhD; Johanna Timonen[1], PhD

School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Corresponding Author:
Maria Sääskilahti, MSc
School of Pharmacy
Faculty of Health Sciences
University of Eastern Finland
PO Box 1627
Kuopio, 70211
Finland
Phone: 358 403552505
Email: maria.saaskilahti@uef.fi

Abstract

Background: Patient portals have been introduced in several countries in the last few decades. Despite worldwide objectives of introducing patient portals, nationwide portals are rare, and studies about their use are limited. Finland is one of the forerunners in developing nationwide health data systems. A nationwide patient portal, My Kanta, for viewing electronic prescriptions and health data has been phased in, starting in 2010.

Objective: The aim of this study was to investigate what functions Finnish pharmacy customers use in My Kanta, their perceptions of the service's usability, and how satisfied users are with My Kanta overall.

Methods: In spring 2019, questionnaires (N=2866) were distributed from 18 pharmacies of varying size across mainland Finland to adult pharmacy customers purchasing prescription medications for themselves or for their child under 18 years of age. Questions were asked about the use and usability of the patient portal by means of structured and Likert-scale questions. Statistical analyses included frequencies, means, medians, chi-square tests, Fisher exact tests, and Kruskal-Wallis tests.

Results: In total, 994 completed questionnaires of 2866 delivered questionnaires (34.68%) were returned. The most-used My Kanta functions were browsing prescription information (781/802, 97.4%), records of health care visits (772/802, 96.3%), and results of laboratory tests and x-ray examinations (722/804, 89.8%). Most users (558/793, 70.4%) had also requested a prescription renewal using the service. My Kanta was perceived as easy to log into (772/816, 94.6%) and clear to view (709/808, 87.7%). Most users considered the service useful for monitoring their health information (753/813, 92.6%) and felt that it provides a good overall picture of the medications prescribed to them (711/813, 87.5%). The majority of users found the information recorded about them easy to understand (684/800, 85.5%). Of the users, 16.7% (135/807) disagreed with the statement that the information they were looking for was easy to find. Approximately two-thirds (501/814, 61.5%) of users did not know whether it is easy to view in which pharmacies and health care units their prescription information has been viewed, and over one-third (306/805, 38.0%) did not know whether it is easy to view in which health care units their health information has been processed. Approximately one-fifth of participants (181/805, 22.5%) feared that unauthorized persons might view their information and that their electronically saved prescription and health information might disappear (180/810, 22.2%). In addition, 16.1% (129/799) expressed interest in receiving guidance on My Kanta use. The vast majority of users (719/804, 89.4%) were satisfied with the service overall.

Conclusions: Pharmacy customers were satisfied with the nationwide patient portal. It was mostly used for browsing e-prescriptions and medical records. Overall, the usability of the service was good. However, users need to be better informed about data privacy and security issues, and guidance on using the portal needs to be improved.

(J Med Internet Res 2021;23(7):e25368) doi:10.2196/25368

KEYWORDS

patient portal; electronic prescription; electronic health records; usability; pharmacy customer; perception; experience; survey

XSL•FO
RenderX

Introduction

Background

Patient portals displaying electronic health records have been developed worldwide in the last few decades. One aim of patient portals is to increase patients' empowerment and their responsibility for their own health and well-being [1-6]. In most countries, for example, in the United States, the Netherlands, and the United Kingdom, portals cater to patients of a specific organization, with a particular disease, or of a specific region [3,5,7-11], while nationwide portals are rare. Nationwide portals have, however, been introduced in the Nordic countries and Estonia [12-18].

The contents and functions of portals vary between countries and portals. Portals based on an organization, disease, or region often include functions such as viewing medical notes, visit summaries, diagnoses, laboratory tests, and medications; scheduling appointments; renewing prescriptions; and secure messaging with health care professionals [3,7-10]. The contents of nationwide portals are broadly the same as in organization-based or disease-specific portals, allowing patients to monitor health and medication data [2,12-15,17,18]. However, nationwide portals seldom include the opportunity to request a prescription renewal, schedule appointments, or communicate with health care professionals. On the other hand, some nationwide portals allow users to declare organ donation testaments, access log lists (a list where one can view who or which organization has viewed one's information), or restrict access to health records. Nationwide patient portals also vary in terms of how widely available the health records are from different units (eg, hospitals, private providers, or health centers) or regions. For example, some private providers in Sweden and some regions in Norway do not provide access to their records, whereas electronic health records from all units in Denmark and Finland are available to patients.

Studies about the use and patients' perceptions of organization-based or disease-specific portals are common, most of them conducted in the United States and the Netherlands. Studies show that the most-used functions of patient portals are viewing laboratory results and medical records [3,5,7-10]. Studies on the usability of services have shown that, overall, patients are satisfied with the services [3,8,10,19,20]. However, there are few studies about the use and usability of nationwide patient portals [13,14,17]. The World Health Organization is encouraging its member states to develop national digital health systems [21]. It is thus important to study the use and usability of nationwide portals to provide countries developing such systems information to support their development work. In addition, users' perceptions of portal use are also important to help make existing systems more useful and user-friendly.

In Finland, a nationwide patient portal, My Kanta, has been introduced. The aims of this study were to investigate what functions are used in My Kanta, users' perceptions of the service's usability, and how satisfied users are with My Kanta overall.

Study Context

In Finland, Kanta services are nationwide digital health care and social welfare services intended to be used by health care professionals, pharmacies, and citizens throughout the country [22]. The services have been phased-in starting in 2010 and are continuously under development. Kanta services are maintained and developed in cooperation with national authors such as the Social Insurance Institution of Finland, the Ministry of Social Affairs and Health, and the National Institute for Health and Welfare [23].

In Finland, it has been obligatory to issue all prescriptions electronically since the beginning of 2017 [24]. Every stage of the Finnish prescribing system is electronic. A physician saves electronic prescriptions (e-prescriptions) to a centralized database, called the Prescription Centre, from where they can be retrieved for dispensing in every pharmacy across Finland. In exceptional situations, such as power blackouts, paper or telephone prescriptions are permitted, but conventional prescriptions are saved to the Prescription Centre at pharmacies when dispensing the prescription for the first time. All Finnish public health care units (primary and special health care) and private health care providers record patients' health care data in the nationwide Patient Data Repository [25]. Health data has been recorded in this service from public health care since 2013 and from private units since 2016. To access the Kanta services, health care and pharmacy professionals need to verify their identity with strong electronic identification (a smart card). All data viewing and processing by professionals can be traced.

Prescriptions and health data recorded in one unit can be shared with other units, with the patient's consent [25]. e-Prescriptions can be viewed in pharmacies and health care units via oral consent from the patient. To share health data, informed consent from the patient has to be approved and saved in Kanta services. The consent is valid until further notice and covers all the health data recorded in the Patient Data Repository. However, the patient can deny the sharing of certain health data (eg, certain health care visits or all the data of certain units) or e-prescriptions. All consents and refusals can be approved or canceled by the patient in health care units or in My Kanta.

My Kanta, a part of Kanta services, is an online service allowing information about e-prescriptions and health data to be viewed by patients [26]. Every person with a Finnish identity number and an ID for electronic services, such as an online banking code, can sign into the service. My Kanta shows an overview of the user's e-prescriptions: when and where the prescription was issued, name of the prescriber, dosage instructions, valid date of prescription, whether there is any medication left, when and where the medication was purchased, and whether the prescription has been renewed. Health data shown in the service consist of records of health care visits, diagnoses, critical risk factors, laboratory tests, x-ray examinations, referrals, health and care plans, and medical certificates and statements (issued, for example, to secure allowances from the Finnish Social Insurance Institute). Using My Kanta, patients can request a prescription renewal, print out a summary of their e-prescriptions, consent to or limit the disclosure of personal data, record living wills and organ donation testaments, and

XSL•FO
RenderX

view in which health care units and pharmacies their personal data has been viewed or processed (later referred to as browsing disclosed information). Guardians can view the health data and e-prescriptions of dependents under 10 years of age and also request a renewal of dependents' prescriptions. This paper focused on functions concerning participants' personal data.

Methods

Data Collection

In spring 2019, a questionnaire survey was conducted among pharmacy customers aged 18 years or older who were purchasing prescription medications for themselves or for their child under 18 years of age. Questionnaires (Multimedia Appendix 1) were distributed by 18 community pharmacies of varying size across mainland Finland (of 623 total pharmacies in Finland). Pharmacies were recruited from all 6 Regional State Administrative Agency areas in mainland Finland. One university pharmacy branch (owned by a university but operating as a privately owned pharmacy); one large, privately owned, urban pharmacy; and one small, privately owned, rural pharmacy were chosen from each region using convenience sampling. Pharmacies were instructed to offer questionnaires to all eligible customers after dispensing prescription medications. The number of questionnaires delivered to a pharmacy was in relation to the number of prescriptions dispensed annually at that pharmacy and varied between 40 and 320 questionnaires. In total, 3560 questionnaires were delivered to the pharmacies. Pharmacists requested customers to complete the questionnaires at home and to post them in return envelopes to the research group. Pharmacies did not keep a record of customers taking questionnaires or refusing to participate in the study. Pharmacies distributed questionnaires for a maximum of 2 weeks. After the study period, pharmacies informed the research group of how many questionnaires remained, to allow the response rate to be calculated. In total, 2866 questionnaires were distributed.

Questionnaire

The questionnaire was designed based on My Kanta pages and previous studies about patient portals [3-5,16,19,26-29]. It was tested for face validity by 3 researchers experienced with designing questionnaires, before a pilot test at a pharmacy. In the pilot test, pharmacy customers completed questionnaires and discussed the questions and their intelligibility with researchers. Minor revisions were made as a result.

The questionnaire included 22 questions and was divided into 3 parts. The first part was for all participants and concerned background information, the second was for users of the service, and the third part was for those who did not use the service. Questions about background information (ie, gender, age, education, region, internet use, internet use for searching health-related information, existence of chronic diseases, and number of currently used, regular prescriptions) were structured except for age and number of currently used, regular prescriptions, which took the form of open-ended questions.

This paper reports results from 4 of the questions from the second part of the questionnaire. Two structured questions concerned the use of different functions in My Kanta, asking "Have you used the following functions in My Kanta?" The first question concerned e-prescriptions and health data, and response options were "often," "sometimes," "rarely," and "never." The second question concerned consenting and limiting consent, for which the response options were "yes," "no," and "do not know." A 5-point Likert-scale question, "What do you think about the following statements?" with response options "fully agree," "agree to some extent," "disagree to some extent," "fully disagree," and "do not know," included 18 statements about the service and its usability. A 6-point Likert-scale question concerned users' overall satisfaction with My Kanta: "How satisfied are you with My Kanta as a whole?" with responses ranging from 1 ("not satisfied at all") to 6 ("very satisfied").

Statistical Analysis

Statistical analyses were conducted using SPSS software (version 25.0; IBM Corp). Descriptive analyses included frequencies, means, and medians. Differences in the use of My Kanta functions between participants were examined using the chi-square test and Fisher exact test. The nonparametric Kruskal-Wallis test was used to analyze differences between means in independent groups for satisfaction with My Kanta. Statistical significance was determined as $P<.05$.

For the analyses, participants' years of birth were converted to ages, in years, and then categorized into 4 age groups: 18-34, 35-59, 60-74, and ≥75. In the questionnaire, education had 5 response options: "basic education," "vocational degree," "secondary school graduate," "lower university degree," and "higher university degree." For the analyses, "vocational degree" and "secondary school graduate" were combined into "secondary education," and "lower university degree" and "higher university degree" were combined into "university degree." In the question concerning existence of chronic diseases, the responses "do not know" were regarded as missing values due to the low number of these responses. The number of regularly used prescription medications was placed into 3 groups: 0, 1-4, and ≥5. In the Likert-scale question about users' perceptions of My Kanta and its usability, response options "fully agree" and "agree to some extent" were combined, and "fully disagree" and "disagree to some extent" were also combined.

Ethical Statement

According to the National Instruction for Research Ethics [30], this study did not require ethical approval. However, approval was obtained at the request of the funding organization from the Committee on Research Ethics of the University of Eastern Finland (number 23/2018). Participation in the study was voluntary; responding to the questionnaire and posting it to the researchers was regarded as informed consent to participate. No incentives were provided for participating in the study. Pharmacy owners permitted the distribution of questionnaires at their pharmacies.

XSL•FO
RenderX

Results

Study Population

In total, 996 questionnaires were returned. Two of them were blank and were therefore excluded from the study. The final study sample comprised 994 responses from the 2866 questionnaires distributed (34.68%). Over two-thirds (687/990, 69.4%) of participants were female (Table 1), and the mean age was 62 years (range 18-99, median 66). Participants were from all 6 regions across Finland. Of all participants, 82.5% (820/994) were My Kanta users. The characteristics of My Kanta users were very similar to those of all participants except for more frequent internet use and internet use for searching health-related information. In addition, there appears to be less participants aged 75 years or older among My Kanta users.

XSL•FO
RenderX

Table 1. Study population characteristics.

Variable	Total participants (N=994)[a], n (%)	My Kanta users (n=820)[a], n (%)
Gender	990	819
Female	687 (69.4)	576 (70.3)
Male	303 (30.6)	243 (29.7)
Age (years)	958	791
18-34	54 (5.6)	50 (6.3)
35-59	269 (28.1)	236 (29.8)
60-74	467 (48.7)	396 (50.1)
≥75	168 (17.5)	109 (13.8)
Education	994	820
Basic education	185 (18.6)	129 (15.7)
Secondary education	523 (52.6)	444 (54.1)
University degree	286 (28.8)	247 (30.1)
Region	992	818
Southern Finland	135 (13.6)	107 (13.1)
Southwestern Finland	144 (14.5)	109 (13.3)
Western and Central Finland	192 (19.4)	155 (18.9)
Eastern Finland	224 (22.6)	189 (23.1)
Northern Finland	222 (22.4)	193 (23.6)
Lapland	75 (7.6)	65 (7.9)
Internet use	987	814
Daily or on several days a week	851 (86.2)	772 (94.8)
Once a week or less often	79 (8.0)	42 (5.2)
Not at all	57 (5.8)	0 (0.0)
Internet use for searching health-related information	991	819
Yes	842 (85.0)	770 (94.0)
No	149 (15.0)	49 (6.0)
Has any chronic disease diagnosed by a physician	982	809
Yes	823 (83.8)	682 (84.3)
No	140 (14.3)	113 (14.0)
Does not know	19 (1.9)	14 (1.7)
Current use of regular prescription medications	942	780
0	101 (10.7)	87 (11.2)
1-4	604 (64.1)	496 (63.6)
≥5	237 (25.2)	197 (25.3)
Use of the My Kanta service	994	820
Yes	820 (82.5)	820 (100.0)
Has used it but is not going to use it anymore	21 (2.1)	0 (0.0)
Has never used it	153 (15.4)	0 (0.0)

[a]Some participants did not answer the question. Therefore, the total for each variable category differs.

XSL•FO
RenderX

Use of Different Functions in My Kanta

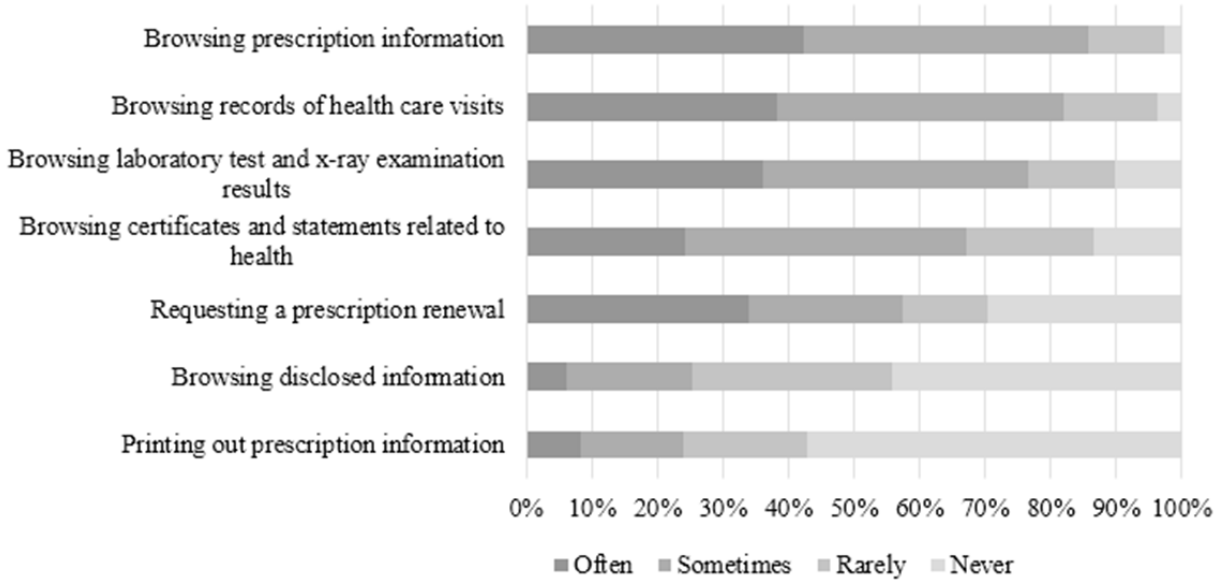
Use of Functions Concerning e-Prescriptions and Health Data

The most-used My Kanta functions concerning e-prescriptions and health data were browsing prescription information (781/802, 97.4%), records of health care visits (772/802, 96.3%), and results of laboratory tests and x-ray examinations (722/804, 89.8%) (Figure 1). Over one-third of users used these functions often (340/802, 42.4%; 306/802, 38.2%; 290/804, 36.1%; respectively). Browsing prescription information was associated with internet use (P<.001); use of the internet to search for health-related information (P<.001); existence of chronic diseases (P<.001); and number of currently used, regular prescription medications (P<.001) (Multimedia Appendix 2). Records of health care visits was also associated with internet use (P=.002); use of the internet to search for health-related information (P<.001); existence of chronic diseases (P=.02); and number of currently used, regular prescription medications (P=.003) . For example, those who seldom used the internet or did not use the internet to search for health-related information more commonly had never browsed prescription information or records of health care visits. Users without chronic diseases were less likely to have often browsed prescription information or records of health care visits. Users who currently used ≥5 regular prescription medications more commonly had often browsed prescription information and records of health care visits.

Of My Kanta users, 70.4% (558/793) had requested a renewal of their prescriptions in the service (Figure 1). Approximately one-third did this often (268/793, 33.8%). Requesting a prescription renewal was associated with education (P=.01), internet use (P=.01), existence of chronic diseases (P<.001), and number of currently used, regular prescription medications (P<.001) (Multimedia Appendix 2). For example, users with a university degree, those who seldom used the internet, users without chronic diseases, or those who did not use any regular prescription medications more commonly had never requested a prescription renewal. Instead, those who currently used ≥5 regular prescription medications more commonly had often requested a prescription renewal in My Kanta.

The least-used functions were browsing disclosed information (438/787, 55.7%) and printing out prescription information (327/765, 42.7%) (Figure 1). Users with a basic education more commonly browsed disclosed information often, whereas users with a university degree more commonly had never browsed disclosed information (P=.001) (Multimedia Appendix 2). Printing out prescription information was associated with gender (P=.02), age (P=.001), existence of chronic diseases (P<.001), and number of currently used, regular prescription medications (P<.001). For example, men less often had never printed out prescription information. Users aged 75 years and older or those who currently used ≥5 regular prescription medications more commonly printed out prescription information often. Users without chronic diseases more commonly had never printed out prescription information.

Figure 1. Frequency of using functions concerning electronic prescriptions and health data in My Kanta.



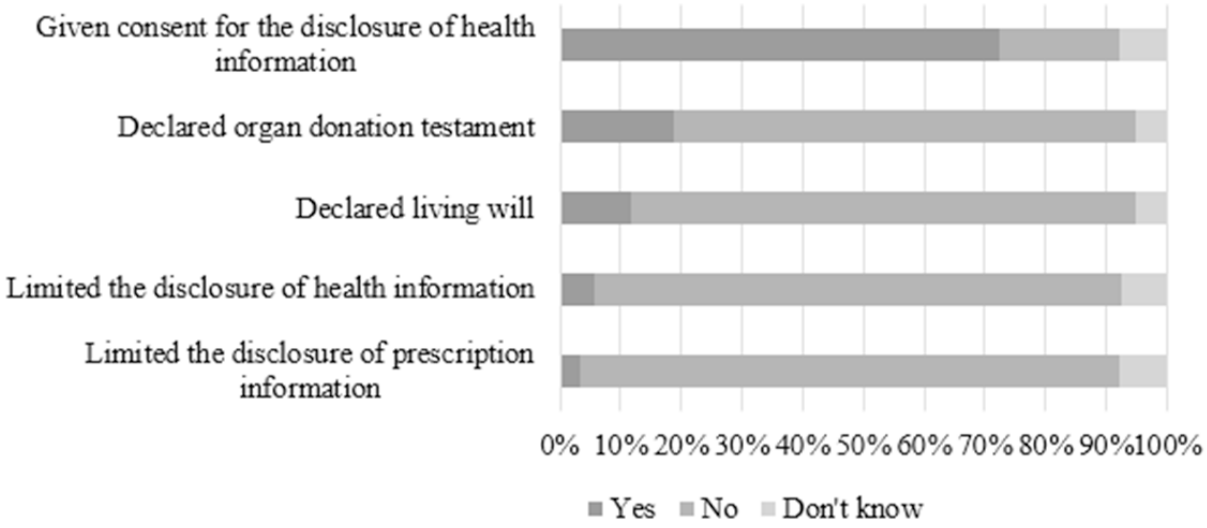
Consents and Limitations

Most My Kanta users (584/808, 72.3%) had consented to disclosure of their health information in the service (Figure 2). Approximately one-fifth of users (151/800, 18.9%) had declared an organ donation testament and approximately one-tenth (95/803, 11.8%) had declared a living will in the service. It was rare that participants limited disclosure of health (45/797, 5.6%) and prescription information (26/799, 3.3%). Giving consent for disclosure of health information was associated with age (P=.001) and internet use (P<.001) (Multimedia Appendix 3). Young participants (18-34 years), more commonly than older participants, did not know whether they had given consent for disclosure of health information. In addition, those who seldom used the internet had less often consented to disclosure of health information in the service. Age was associated with limiting

XSL•FO
RenderX

the disclosure of health (*P*=.02) and prescription information (*P*=.04). For example, young participants (18-34 years), more commonly than older participants, did not know whether they had limited the disclosure of health data. Declaring an organ donation testament was associated with age (*P*<.001), education (*P*<.001), and internet use (*P*=.03). For example, participants aged 18-59 years had more commonly declared an organ donation testament, whereas those 75 years and older had declared it less often. In addition, those with a basic education had declared an organ donation testament less often, while the declaration was more common among participants with a university degree.

Figure 2. Use of functions concerning consents and limitations in My Kanta.



Usability of My Kanta

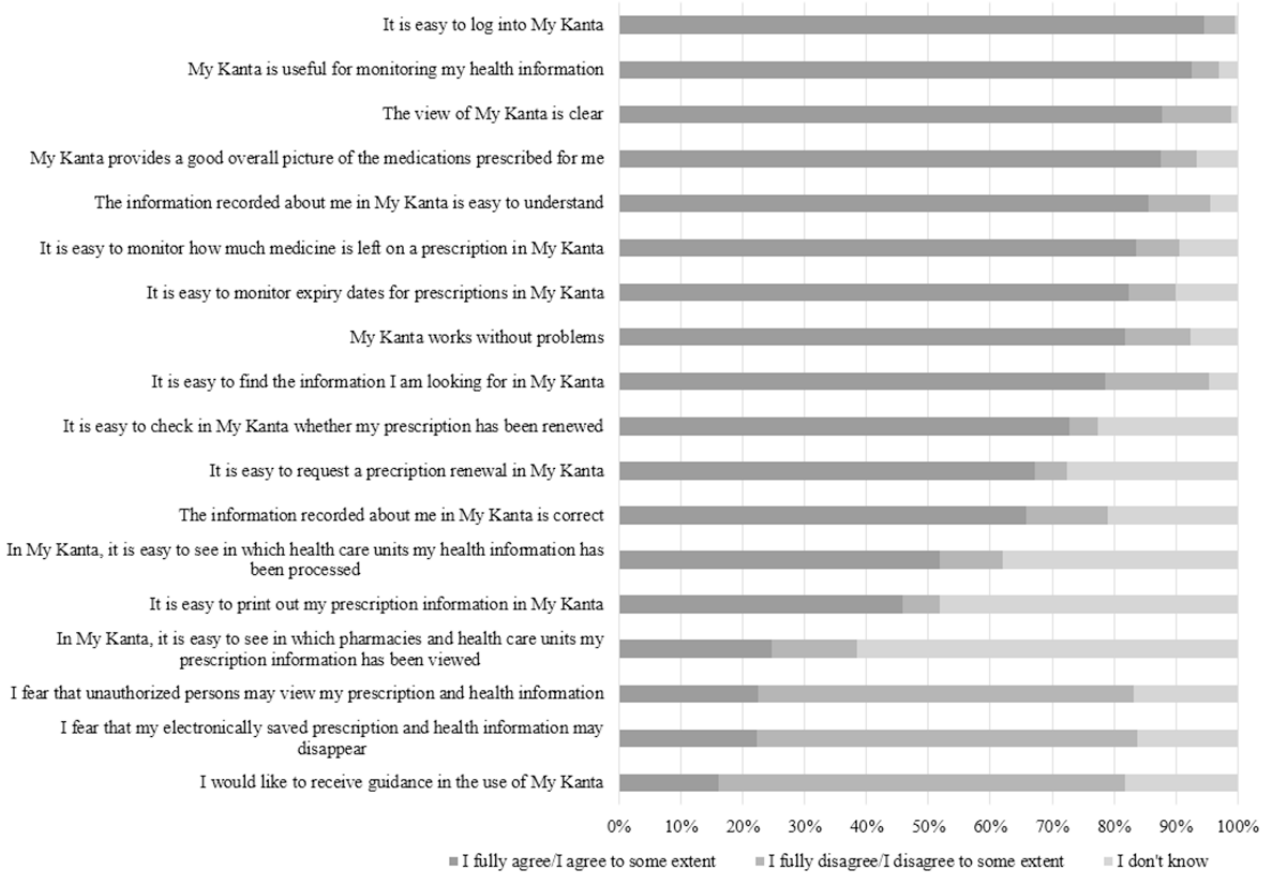
Most My Kanta users reported that My Kanta is easy to log into (772/816, 94.6%), its view is clear (709/808, 87.7%), it works without problems (655/802, 81.7%), and it is easy to find the information they are looking for (634/807, 78.6%) (Figure 3).

Most users said My Kanta is useful for monitoring their health information (753/813, 92.6%) and that the service provides a good overall picture of the medications prescribed to them (711/813, 87.5%) (Figure 3). A high proportion of users agreed that, in My Kanta, it is easy to monitor how much medicine is left for a prescription (682/817, 83.5%) and what the expiry dates for prescriptions are (668/811, 82.4%). Most users found the information recorded about them easy to understand (684/800, 85.5%).

Conversely, 16.7% (135/807) of users disagreed with the statement that it is easy to find the information they are looking for (Figure 3). Approximately two-thirds (501/814, 61.5%) of users did not know whether it is easy to view in which pharmacies and health care units their prescription information has been viewed. In addition, over one-third (306/805, 38.0%) did not know whether it is easy to view in which health care units their health information has been processed. Approximately one-fifth of participants feared that unauthorized persons might view their prescription and health information (181/805, 22.5%) and that their electronically saved prescription and health information might disappear (180/810, 22.2%). Altogether, 16.1% (129/799) of My Kanta users expressed interest in receiving guidance on using the service.

XSL•FO
RenderX

Figure 3. Participants' perceptions of My Kanta and its usability.



Overall Satisfaction With My Kanta

On the 6-point Likert scale, 89.4% (719/804) of My Kanta users had rated their overall satisfaction with the service within a range of 4-6 (mean 4.8; median 5) (Table 2). Overall satisfaction differed significantly across participants' education levels (P=.03) and frequency of internet use (P=.001). Participants with a basic education were more satisfied with My Kanta than those with a university degree, and participants who used the internet daily or several times a week were more satisfied than those who used the internet once a week or less.

XSL•FO
RenderX

Table 2. Participants' overall satisfaction rating[a] with My Kanta (N=804).

Participant group	Rating						*P* value
	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	6, n (%)	
All participants	7 (0.9)	16 (2.0)	62 (7.7)	153 (19.0)	373 (46.4)	193 (24.0)	N/A[b]
Gender							.11
Female	5 (0.9)	13 (2.3)	42 (7.4)	118 (20.9)	256 (45.4)	130 (23.0)	
Male	2 (0.8)	3 (1.3)	20 (8.4)	35 (14.6)	116 (48.5)	63 (26.4)	
Age (years)							.33
18-34	0 (0.0)	0 (0.0)	0 (0.0)	10 (20.0)	29 (58.0)	11 (22.0)	
35-59	0 (0.0)	2 (0.9)	16 (6.9)	51 (22.0)	105 (45.3)	58 (25.0)	
60-74	4 (1.0)	11 (2.8)	35 (9.0)	64 (16.5)	187 (48.1)	88 (22.6)	
≥75	3 (2.9)	3 (2.9)	10 (9.6)	23 (22.1)	41 (39.4)	24 (23.1)	
Education							.03
Basic education	3 (2.3)	2 (1.6)	5 (3.9)	21 (16.4)	54 (42.2)	43 (33.6)	
Secondary education	3 (0.7)	9 (2.1)	36 (8.3)	79 (18.3)	205 (47.5)	100 (23.1)	
University degree	1 (0.4)	5 (2.0)	21 (8.6)	53 (21.7)	114 (46.7)	50 (20.5)	
Internet use							.001
Daily or on several days a week	5 (0.7)	12 (1.6)	55 (7.2)	142 (18.7)	357 (47.0)	188 (24.8)	
Once a week or less often	1 (2.6)	2 (5.1)	7 (17.9)	10 (25.6)	14 (35.9)	5 (12.8)	
Internet use for searching health-related information							.06
Yes	5 (0.7)	15 (2.0)	56 (7.4)	142 (18.8)	354 (46.8)	184 (24.3)	
No	2 (4.3)	1 (2.1)	6 (12.8)	11 (23.4)	18 (38.3)	9 (19.1)	
Has any chronic disease diagnosed by a physician							.18
Yes	5 (0.7)	16 (2.4)	53 (7.9)	131 (19.6)	306 (45.9)	156 (23.4)	
No	2 (1.8)	0 (0.0)	8 (7.1)	19 (17.0)	51 (45.5)	32 (28.6)	
Current use of regular prescription medications							.54
0	1 (1.2)	1 (1.2)	6 (7.0)	25 (29.1)	33 (38.4)	20 (23.3)	
1-4	4 (0.8)	11 (2.3)	37 (7.6)	85 (17.5)	231 (47.5)	118 (24.3)	
≥5	1 (0.5)	3 (1.6)	15 (7.8)	36 (18.8)	92 (47.9)	45 (23.4)	

[a]Rating scale: 1 ("not satisfied at all") to 6 ("very satisfied").

[b]N/A: not applicable.

Discussion

Principal Results and Comparison With Prior Work

Finnish pharmacy customers mostly used the nationwide patient portal to browse their e-prescriptions and records of health care visits. e-Prescriptions and My Kanta were introduced in Finland in 2010. Since 2017, all prescriptions have been issued electronically, and My Kanta is the only place where patients themselves can find up-to-date information about their e-prescriptions. The frequent browsing of e-prescriptions is therefore understandable. However, the frequent use of this function is a positive finding as, in a study conducted in 2015 in Finland, pharmacy customers felt that the biggest problem with e-prescriptions was keeping up to date with their medication [31]. In this study, users reported that the patient portal provides a good overall picture of their medication and makes e-prescriptions easy to monitor. These findings, and the fact that portal use has significantly increased in the last few years [32], suggest that people have learned to monitor their prescription information via the online service. Use of the patient portal mainly to view prescriptions differs from the findings of previous studies about patient portals [9,10,13,14,17]. This may be because a fully electronic, nationwide prescribing system integrated into a nationwide electronic health record system is rare worldwide [33], and the use of patient portals to monitor medication is not as essential in other countries.

Although monitoring e-prescriptions via My Kanta has been regarded as easy, almost half of My Kanta users, especially older users and those using several regular prescription medications, have printed out prescription information via the service. This suggests that although most My Kanta users can

XSL•FO
RenderX

manage their medication via the online service, for others, the printed information is still necessary.

In addition to e-prescriptions, almost all My Kanta users browsed records of their health care visits and test results via the service. My Kanta was perceived as useful for monitoring health data. This is in line with previous studies, which found viewing medical records and results of laboratory tests to be the most-used or useful functions in patient portals [3,5,7,10,13,14]. In previous studies, health data have mainly been used to prepare for health care visits, to reread medical information after visits, and to become more aware and involved in patients' own health and care. Future studies should examine the reasons for My Kanta use, in order to find out what role the patient portal has in patients' involvement in their own health and care.

This study showed no differences in the frequencies of using My Kanta for monitoring health and prescription information between user characteristics (ie, gender, age, and education). Instead, using the internet only seldom and not using the internet to search for health-related information were associated with not using My Kanta for browsing health and prescription information. This suggests that information in patient portals is browsed by those who are generally interested in their own health information and search for it on the internet. In agreement with these results, in a previous study, greater health literacy was associated with the use of a patient portal to check test results, whereas gender, age, and education were not [4].

The vast majority of users had used the service to request a prescription renewal. This function has been available in My Kanta since 2015. According to Kanta services' statistical reports, the number of prescription renewal requests is continuously increasing [32]. In 2019, approximately 250,000 renewal requests were submitted monthly via My Kanta (of approximately 2.5 million monthly issued e-prescriptions in Finland). A study conducted in Finland in 2019 showed that physicians regard the fact that patients can request a prescription renewal in My Kanta as largely beneficial, and one reason was that it saves nurses time [34]. However, physicians also saw this function as problematic, as patients can send a request for any medications to any health care units across Finland regardless of where the prescription was issued. Physicians also felt that allowing patients to submit their own renewal requests may cause difficulties in pharmacotherapy monitoring, as physicians have to search for all necessary information in support of a renewal. In the future, it will be important to study how patients' renewal requests have affected physicians' workloads and pharmacotherapy monitoring and, therefore, medication safety. The opportunity to request a prescription renewal is rare in nationwide patient portals. To the best of our knowledge, the function is available in Denmark and Iceland [2,12], although there are no studies reporting the use of this function.

Declaring an organ donation testament or a living will was a rarely used function in My Kanta. Of countries with a nationwide patient portal, at least in Denmark, Estonia, and Iceland, patients have an opportunity to register an organ donor testament [12,18], but studies on the use of this function are unavailable. The purpose of an organ donation testament or a living will is to help health care professionals and patient's relatives make decisions relating to care in unexpected situations [35]. In Finland, according to law, the organs of a deceased person can be salvaged to treat other patients unless the deceased had previously declined [36]. My Kanta is an easy way to record an organ donation testament and living will, and, via the service, wills are secure and available for health care professionals in situations where patients are not able to express their will themselves. The significance and importance of expressing one's will should be clarified for citizens, to increase the use of these functions. This study showed that younger people and people educated with a university degree were more likely to declare an organ donation testament.

According to this study, pharmacy customers have rarely limited the disclosure of e-prescriptions and health data. In Finland, one key aim of e-prescriptions is to improve the management of overall medication [24]. It is important that health care professionals can observe patients' overall medication whenever needed. Another aim of Kanta services is to enable cooperation between health care units and secure the continuity of care [22]. The findings of this study suggest that patients have not prevented achievement of these aims by limiting data disclosure. However, in previous studies, physicians found it difficult to view patients' overall medication via the Prescription Centre, as there is no list of currently used medications [34,37]. This problem will be solved in the future, as Kanta services are developing a national medication list where up-to-date information about currently used medications is available [38]. Further studies are needed to investigate how health care professionals experience the usability of shared e-prescriptions and health data when caring for patients.

Pharmacy customers were, overall, satisfied with My Kanta. This is in line with previous studies [3,13,17,39], which showed that users are largely satisfied with the patient portals they are using. Compared to a study conducted in 2015 that investigated viewing e-prescriptions via My Kanta [16], pharmacy customers' perceptions about the usability of the service have remained broadly the same. This is encouraging, as My Kanta is perceived as easy to sign into and monitor e-prescriptions information, the service works without problems, and its layout is regarded as clear. However, there are also some challenges involved in the use of My Kanta that have remained. For example, a substantial proportion of users still did not know whether it is easy to view in which health care units or pharmacies their information has been viewed and processed. This may indicate that users do not know that this information can be found in the service or that they have not tried to search for it. These assumptions are supported by the finding that almost half of users had never browsed disclosed information in My Kanta. However, some My Kanta users were worried that unauthorized persons might view their information. The situation has not changed since 2015 when the issue was last studied [40]. It is therefore important to inform people and My Kanta users, specifically, about the data protection and privacy procedures in Kanta services [41] and that disclosed information can be checked in My Kanta. This might ease unnecessary concerns about data protection.

XSL•FO
RenderX

Data recorded in My Kanta were mostly perceived as easy to understand. However, approximately one-tenth of users disagreed. This is supported by some previous studies, which revealed that the language used in patient portals is sometimes too complicated for patients and should be simplified [10,17]. In Finland, the patient portal is meant for all citizens. Health care professionals therefore need to pay attention to the language they use when recording health data in the Patient Data Repository. The aim of a patient portal is to make patients more involved in their own health and well-being, but this will not be achieved if patients do not understand the information provided in the service. In addition to simple language, the patient portal has to be simple enough for everyone to use. Of My Kanta users, approximately 16% expressed the desire to receive guidance on My Kanta use. This means that current information and guidance about My Kanta (eg, My Kanta pages, frequently-asked-question pages, and the online course) [26,42,43] need to be improved. As all nationally organized guidance is available only on the internet, face-to-face guidance about the use of My Kanta might also be needed. According to our previous study, main reasons for nonuse of My Kanta were the lack of need and tools [44]. In addition, some pharmacy customers had difficulties with My Kanta use and some were unfamiliar with the service. These results also underline the need for improving the guidance and information about the service.

Strengths and Limitations

This study had several strengths. The patient portal studied here is nationwide and available to everyone living in Finland with internet access and an ID for electronic services. By distributing questionnaires at pharmacies to customers purchasing prescription medications, we reached a target population likely to need to use the patient portal. The study sample was large and included participants across Finland. We achieved our goal of reaching both users and nonusers of the patient portal, and these were distributed similarly by background information except for internet-related characteristics. The questionnaire did not include validated measures, but it was designed based on previous studies about patient portals [3-5,16,19,27-29,45], and some of the questions reported in this paper were based on previous surveys but had minor revisions [16,19,28,45]. The response rates for the questions reported in this paper were high (93%-100%), and, thus, it can be assumed that the questions were understandable.

This study also had limitations. Of the 3560 questionnaires delivered to the pharmacies, 694 questionnaires were not distributed. Most of the pharmacies distributed all the questionnaires, but a few pharmacies distributed one-half or less of the questionnaires delivered to them. We lacked the information on whether these pharmacies attempted to distribute these remaining questionnaires or if the pharmacies were not motivated to do so. The survey response rate was low. We had no information about who declined to participate in the study. As a result, the response rate may be even lower than reported. It was also lower than in studies conducted with the same method earlier in Finland (40%-44%) [16,46]. The trend in survey response rates has generally been declining in recent decades [16,46-49]. We do not have comparable statistics about pharmacy customers purchasing prescription medications in Finland, but, compared to customers receiving reimbursement for medication costs under the Health Insurance Scheme [50], the participants were older and more often women. This is also in line with trends in previous survey studies [16,46-48]. However, participants' characteristics were similar to those in studies conducted previously with the same method [16,46].

Conclusions

Pharmacy customers were satisfied with the nationwide patient portal. It was mostly used for browsing e-prescriptions and medical records. The usability of the service was mainly good, but users need to be better informed about data privacy and security issues as well as the guidance available for use of the portal.

Acknowledgments

The authors are grateful to the pharmacies that distributed the questionnaire. The study was funded by The Social Insurance Institution of Finland (Kela). The funding organization had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The opinions expressed in this document are those of the authors and do not reflect the official position of Kela.

Authors' Contributions

MS, RA, and JT designed the study and collected the data. MS conducted the data analyses and drafted the manuscript. All authors participated in discussing the analyses and findings, critically revising the manuscript, and reading and approving the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A questionnaire.
[DOC File , 143 KB - jmir_v23i7e25368_app1.doc]

XSL•FO
RenderX

Multimedia Appendix 2
Frequency of using My Kanta functions concerning electronic prescriptions and health data, and differences between groups.
[DOC File , 139 KB - jmir_v23i7e25368_app2.doc]

Multimedia Appendix 3
Use of My Kanta functions concerning consents and limitations, and differences between groups.
[DOC File , 97 KB - jmir_v23i7e25368_app3.doc]

References

1. Patient engagement. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252269/9789241511629-eng.pdf [accessed 2021-07-01]
2. Background. Sundhed.dk. URL: https://www.sundhed.dk/borger/service/om-sundheddk/om-organisationen/ehealth-in-denmark/background/ [accessed 2021-06-24]
3. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663. [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]
4. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
5. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
6. Information to support well-being and service renewal. eHealth and eSocial Strategy 2020. Ministry of Social Affairs and Health. URL: http://julkaisut.valtioneuvosto.fi/handle/10024/74459 [accessed 2021-07-01]
7. Elston Lafata J, Miller CA, Shires DA, Dyer K, Ratliff SM, Schreiber M. Patients' adoption of and feature access within electronic patient portals. Am J Manag Care 2018 Nov 01;24(11):e352-e357 [FREE Full text] [Medline: 30452203]
8. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' Perceptions of Portal Use Across Care Settings: Qualitative Study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
9. Tsai R, Bell EJ, Woo H, Baldwin K, Pfeffer MA. How Patients Use a Patient Portal: An Institutional Case Study of Demographics and Usage Patterns. Appl Clin Inform 2019 Jan;10(1):96-102 [FREE Full text] [doi: 10.1055/s-0038-1677528] [Medline: 30727003]
10. Vreugdenhil MMT, Ranke S, de Man Y, Haan MM, Kool RB. Patient and Health Care Provider Experiences With a Recently Introduced Patient Portal in an Academic Hospital in the Netherlands: Mixed Methods Study. J Med Internet Res 2019 Aug 20;21(8):13743. [doi: 10.2196/13743] [Medline: 31432782]
11. Shah SGS, Fitton R, Hannan A, Fisher B, Young T, Barnett J. Accessing personal medical records online: a means to what ends? Int J Med Inform 2015 Feb;84(2):111-118 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.005] [Medline: 25453275]
12. Hardardottir GA, Thoroddsen A. National eHealth Implementation: Country Experience. Stud Health Technol Inform 2016;225:168-172. [Medline: 27332184]
13. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden. J Med Internet Res 2018 Nov 01;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
14. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the Level of Electronic Health Literacy Between Users and Nonusers of Digital Health Services: An Exploratory Survey of a Group of Medical Outpatients. Interact J Med Res 2019 Apr 05;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
15. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]
16. Lämsä E, Timonen J, Mäntyselkä P, Ahonen R. Pharmacy customers' experiences with the national online service for viewing electronic prescriptions in Finland. Int J Med Inform 2017 Jan;97:221-228. [doi: 10.1016/j.ijmedinf.2016.10.014] [Medline: 27919380]
17. Zanaboni P, Kummervold PE, Sørensen T, Johansen MA. Patient Use and Experience With Online Access to Electronic Health Records in Norway: Results From an Online Survey. J Med Internet Res 2020 Feb 07;22(2):e16144 [FREE Full text] [doi: 10.2196/16144] [Medline: 32031538]
18. Nøhr C, Parv L, Kink P, Cummings E, Almond H, Nørgaard JR, et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia. BMC Health Serv Res 2017 Aug 07;17(1):534 [FREE Full text] [doi: 10.1186/s12913-017-2482-y] [Medline: 28784173]

XSL•FO
RenderX

19. van der Vaart R, Drossaert CHC, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15:102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
20. Zide M, Caswell K, Peterson E, Aberle DR, Bui AA, Arnold CW. Consumers' Patient Portal Preferences and Health Literacy: A Survey Using Crowdsourcing. JMIR Res Protoc 2016 Jun 08;5(2):e104. [doi: 10.2196/resprot.5122]
21. Global Strategy on Digital Health 2020-2025. World Health Organization. 2021. URL: https://cdn.who.int/media/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_75 [accessed 2021-07-01]
22. What are the Kanta Services? Kanta Services. URL: https://www.kanta.fi/en/web/guest/what-are-kanta-services [accessed 2019-09-20]
23. Partners. Kanta Services. URL: https://www.kanta.fi/en/partners [accessed 2019-10-03]
24. The Act on Electronic Prescription 61/2007. Finlex. 2007. URL: https://www.finlex.fi/fi/laki/ajantasa/2007/20070061 [accessed 2020-09-20]
25. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. FinJeHeW 2018 Dec 04;10(4):381-395. [doi: 10.23996/fjhw.74511]
26. My Kanta Pages. Kanta Services. URL: https://www.kanta.fi/en/my-kanta-pages [accessed 2019-09-20]
27. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
28. Hyppönen H, Hyry J, Valta K, Ahlgren S. Electronic Services in the Social Welfare and Health Care Sector. Citizens' Experiences and Development Needs. National Institute for Health and Welfare (THL). Report 33/2014. 2014. URL: http://www.julkari.fi/handle/10024/125597 [accessed 2019-10-30]
29. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
30. Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review. National Advisory Board on Research Ethics. URL: https://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf [accessed 2020-07-02]
31. Sääskilahti M, Ahonen R, Lämsä E, Timonen J. Benefits and problems of electronic prescriptions - a survey among pharmacy customers in Finland. DOSIS 2016;32(2):129-141.
32. Jormanainen V, Parhiala K, Niemi A, Erhola M, Keskimäki I, Kaila M. Half of the Finnish population accessed their own data: comprehensive access to personal health information online is a corner-stone of digital revolution in Finnish health and social care. FinJeHeW 2019 Nov 02;11(4):298-310. [doi: 10.23996/fjhw.83323]
33. Bruthans J. The state of national electronic prescription systems in the EU in 2018 with special consideration given to interoperability issues. Int J Med Inform 2020 Sep;141:104205. [doi: 10.1016/j.ijmedinf.2020.104205] [Medline: 32492586]
34. Korhonen P, Timonen J, Kauppinen H, Mäntyselkä P, Ahonen R. Electronic prescription renewal procedures - an interview study among Finnish primary care physicians. Lääkärilehti 2019;74(34):1817-1822.
35. Save a living will and organ donation testament in My Kanta Pages. Blog. Kanta Services. URL: https://www.kanta.fi/blogi/-/asset_publisher/1QjC602jKPR6/content/omakantaan-voi-tallentaa-hoito-ja-elinluovutustahdon [accessed 2020-06-18]
36. Act on the Medical Use of Human Organs and Tissues 101/2001. Finlex. 2001. URL: https://www.finlex.fi/fi/laki/ajantasa/2001/20010101 [accessed 2020-05-14]
37. Kauppinen H, Ahonen R, Mäntyselkä P, Timonen J. Medication safety and the usability of electronic prescribing as perceived by physicians - A semistructured interview among primary health care physicians in Finland. J Eval Clin Pract 2017 Dec;23(6):1187-1194. [doi: 10.1111/jep.12759] [Medline: 28474487]
38. Priorities of the Kanta Services. Kanta Services. URL: https://www.kanta.fi/en/web/guest/professionals/priorities-of-the-kanta-services [accessed 2020-06-22]
39. Wass S, Vimarlund V. The Role of PAEHRs in Patient Involvement. J Med Syst 2018 Sep 25;42(11):210 [FREE Full text] [doi: 10.1007/s10916-018-1070-y] [Medline: 30255404]
40. Rattay K, Kauppinen H, Ahonen R, Timonen J. Opinions on the Data Protection of Electronic Prescriptions - a Survey Among Pharmacy Customers in Finland. DOSIS 2018(4):264-275.
41. Data protection and privacy. Kanta Services. URL: https://www.kanta.fi/en/data-protection [accessed 2020-07-02]
42. Get to know your My Kanta Pages. Kanta Services. URL: https://www.kanta.fi/omakanta-verkkokoulu/#/?_k=oqbbyy [accessed 2019-11-13]
43. Frequently asked questions. Kanta Services. URL: https://www.kanta.fi/en/faq [accessed 2020-07-02]
44. Sääskilahti M, Aarnio E, Lämsä E, Ahonen R, Timonen J. Use and non-use of a nationwide patient portal - a survey among pharmacy customers. Journal of Pharmaceutical Health Services Research 2020;11(4):335-342. [doi: 10.1111/jphs.12368]
45. Hyppönen H, Pentala-Nikulainen O, Aalto AM. Electronic services in the social welfare and health care sector 2017. Citizens' experiences and needs. National Institute for Health and Welfare (THL). Report 3/2018. 2018. URL: http://www.julkari.fi/handle/10024/136258 [accessed 2019-10-30]

XSL•FO
RenderX

46. Nokelainen H, Lämsä E, Ahonen R, Timonen J. Reasons for allowing and refusing generic substitution and factors determining the choice of an interchangeable prescription medicine: a survey among pharmacy customers in Finland. BMC Health Serv Res 2020 Feb 03;20(1):82 [FREE Full text] [doi: 10.1186/s12913-020-4894-3] [Medline: 32013951]
47. Tolonen H, Helakorpi S, Talala K, Helasoja V, Martelin T, Prättälä R. 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey. Eur J Epidemiol 2006;21(6):409-415. [doi: 10.1007/s10654-006-9019-8] [Medline: 16804763]
48. Sarnola K, Hämeen-Anttila K, Jyrkkä J. Medicines Barometer 2019. Description of the data and data collection of the population survey. Serial Publication Fimea Develops, Assesses and Informs 8/2019. Kuopio: Finnish Medicines Agency (Fimea); 2019.
49. Heikkilä R, Mäntyselkä P, Hartikainen-Herranen K, Ahonen R. Customers' and physicians' opinions of and experiences with generic substitution during the first year in Finland. Health Policy 2007 Aug;82(3):366-374. [doi: 10.1016/j.healthpol.2006.10.006] [Medline: 17141355]
50. Finnish statistics on medicines 2017. Finnish Medicines Agency and Social Insurance Institution. 2018. URL: https://www.fimea.fi/documents/160140/763020/Suomen_l%C3%A4%C3%A4ketilasto_2017_korjattu_3_painos.pdf/5c50148d-f586-182a-f266-90fa38d8c4dd [accessed 2021-07-01]

Abbreviations

e-prescription: electronic prescription

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by J Bruthans, S MacEwan; comments to author 27.11.20; revised version received 10.12.20; accepted 30.05.21; published 16.07.21.

Please cite as:
Sääskilahti M, Ahonen R, Timonen J
Pharmacy Customers' Experiences of Use, Usability, and Satisfaction of a Nationwide Patient Portal: Survey Study
J Med Internet Res 2021;23(7):e25368
URL: https://www.jmir.org/2021/7/e25368
doi:10.2196/25368
PMID:34269687

©Maria Sääskilahti, Riitta Ahonen, Johanna Timonen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Functionalities and Issues in the Implementation of Personal Health Records: Systematic Review

Nabila Clydea Harahap[1], SKom, MKom; Putu Wuri Handayani[1], SKom, MSc, PhD; Achmad Nizar Hidayanto[1], SKom, MKom, PhD

Faculty of Computer Science, University of Indonesia, Depok, Indonesia

Corresponding Author:
Nabila Clydea Harahap, SKom, MKom
Faculty of Computer Science
University of Indonesia
Kampus UI Depok, Pondok Cina, Beji
Depok, 16424
Indonesia
Phone: 62 85716526799
Email: nabila.clydea@ui.ac.id

Abstract

Background: Functionalities of personal health record (PHR) are evolving, and continued discussions about PHR functionalities need to be performed to keep it up-to-date. Technological issues such as nonfunctional requirements should also be discussed in the implementation of PHR.

Objective: This study systematically reviewed the main functionalities and issues in implementing the PHR.

Methods: This systematic review was conducted using Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines. The search is performed using the online databases Scopus, ScienceDirect, IEEE, MEDLINE, CINAHL, and PubMed for English journal articles and conference proceedings published between 2015 and 2020.

Results: A total of 105 articles were selected in the review. Seven function categories were identified in this review, which is grouped into basic and advanced functions. Health records and administrative records were grouped into basic functions. Medication management, communication, appointment management, education, and self-health monitoring were grouped into advanced functions. The issues found in this study include interoperability, security and privacy, usability, data quality, and personalization.

Conclusions: In addition to PHR basic and advanced functions, other supporting functionalities may also need to be developed based on the issues identified in this study. This paper provides an integrated PHR architectural model that describes the functional requirements and data sources of PHRs.

(J Med Internet Res 2021;23(7):e26236) doi:10.2196/26236

KEYWORDS

personal health record; systematic review; functionalities; issues

Introduction

In health emergencies such as epidemics, natural disasters, or artificial disasters, access to reliable health information becomes crucial for the community [1,2]. As of 2020, the COVID-19 pandemic throughout the world has led to an increasing need for electronic health records (EHRs) to provide reliable health information [3,4]. According to the World Health Organization, the EHR that collects data from various health service providers will provide better patient care during a pandemic, such as preventing and detecting an outbreak [5]. The EHR's function will be more optimal if patients can share their health data with health care providers [6]. Personal health records (PHRs) can help patients share their data with health care providers and provide useful information during health emergencies [2].

The EHR aims to collect health data managed by health care providers, while the PHR aims to collect health data entered by individuals [1]. The PHR was developed with a patient-centered approach in the capture and storage of information [7]. In its simplest form, a PHR is a stand-alone application that is not connected to other systems. Users can access their PHR using commercially available applications to record and analyze daily activities and habits to maintain a healthy lifestyle. In a more complex form, the PHR's health information is connected to

XSL•FO
RenderX

the EHR of the health care provider (tethered PHR) or to various health service data sources (integrated or interconnected PHR). A PHR integrated with an EHR, either through tethering or interconnectivity, provides far more significant benefits than a stand-alone PHR [1].

One of the important PHR research areas is PHR functionality [8]. Previous studies have provided data types and functionalities of PHRs [9] and a guide to evaluate PHR functionalities [10]. Some studies reviewed PHRs used for chronic diseases, which include discussions about their functionalities [11,12]. However, these studies focused only on PHRs in the United States and developed countries [9,11,12]. Moreover, previous studies have also discussed technological issues in implementing or using PHRs, such as data quality [13], personalization [14], privacy [13,15], and usability [14]. These studies still have no clear explanations about how these issues can be included as requirements in implementing PHRs.

Functions or features of PHRs are evolving [9,12], so continued discussions about PHR functionalities need to be held to keep the research up-to-date. In addition, technological issues as nonfunctional requirements [16] in the design and development of a system must be discussed. Technological issues can be defined as constraints and qualities related to the technology used to perform the function [17]. Thus, this paper aims to review the PHR studies focusing on the functionalities and technological issues in building the PHR system. This paper addresses the following research question: What are the main functionalities and issues in the implementation of PHRs? This study can provide PHR design or implementation recommendations to health care management, application developers, policymakers, or other related stakeholders.

Methods

This systematic review was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [18]. PRISMA is suitable for studies related to health care interventions, and it focuses on ways in which authors can ensure the transparent and complete reporting of systematic reviews [19]. The PRISMA checklist for this study is provided in Multimedia Appendix 1.

Search Strategy

The search is conducted using the online databases Scopus, ScienceDirect, IEEE Xplore, MEDLINE, CINAHL, and PubMed. Terms or keywords used to search the articles: (“phr” OR “personal health record” OR “personal medical record” OR “personal health information” OR “personally controlled electronic health record” OR “pcehr” OR “patient portal”) AND (“functionality” OR “features” OR “issues” OR “implementation”). The search was conducted for journal articles and conference proceedings published between January 2015 and December 2020 to ensure that the data were current because the functions of PHRs are evolving.

Eligibility Criteria

The authors defined inclusion criteria as the review guidelines for study selection. The articles included for this study must have full text available and written in English, be original research articles, focus on discussing the electronic PHR platform, and discuss functionalities and/or issues in the implementation of PHRs.

In this study, the PHRs discussed are all PHR types (stand-alone, tethered, and integrated) that provide access to health information or records to patients electronically. Therefore, papers with related terms such as patient health records or patient portals are also included in this review. The authors also reviewed PHRs at the design stage to include conceptual papers in this review.

Study Selection

The study selection consists of the following phases:

1. Keyword or search string was searched in each online database previously mentioned. Duplicated records were checked and removed.
2. The title and abstract of identified articles were selected based on the eligibility criteria. Articles that did not meet inclusion criteria were eliminated.
3. Articles that were not eliminated in the previous stage were read in full text to determine whether they should be included in the review based on the eligibility criteria. Reference lists of the included studies were also checked to identify additional relevant articles.

The first author screened the titles and abstracts based on the eligibility criteria. The same author reviewed full-text versions of the articles that were not excluded from the previous screening. The first author extracted data from selected studies and the second author reviewed the extracted data. Disagreements between the two authors were resolved through discussion. If an agreement could not be made, the third author would determine the decision. We were unable to consistently evaluate the risk of bias due to the variety of methodologies within the studies.

Data Items and Synthesis

Data collection was performed manually using a data extraction form. Information extracted from each article consists of characteristics of selected articles, such as study location, PHR purpose, and methodology, and functionalities of PHRs and issues in PHR implementation

Authors categorized functionalities of PHRs based on their purpose as defined in Bouayad et al [9], Price et al [12], and Genitsaridi et al [10]. For each function category, the authors explained subfunctions or data elements that were implemented or recommended from the selected articles. Moreover, each function category was grouped based on basic and advanced functions defined by Detmer et al [20]. Basic functionalities help people collect, organize, and store health information, while advanced functionalities enable patients to play a more active role in their health [20]. The authors explained PHR implementation issues that are mentioned explicitly or implicitly from the selected articles.

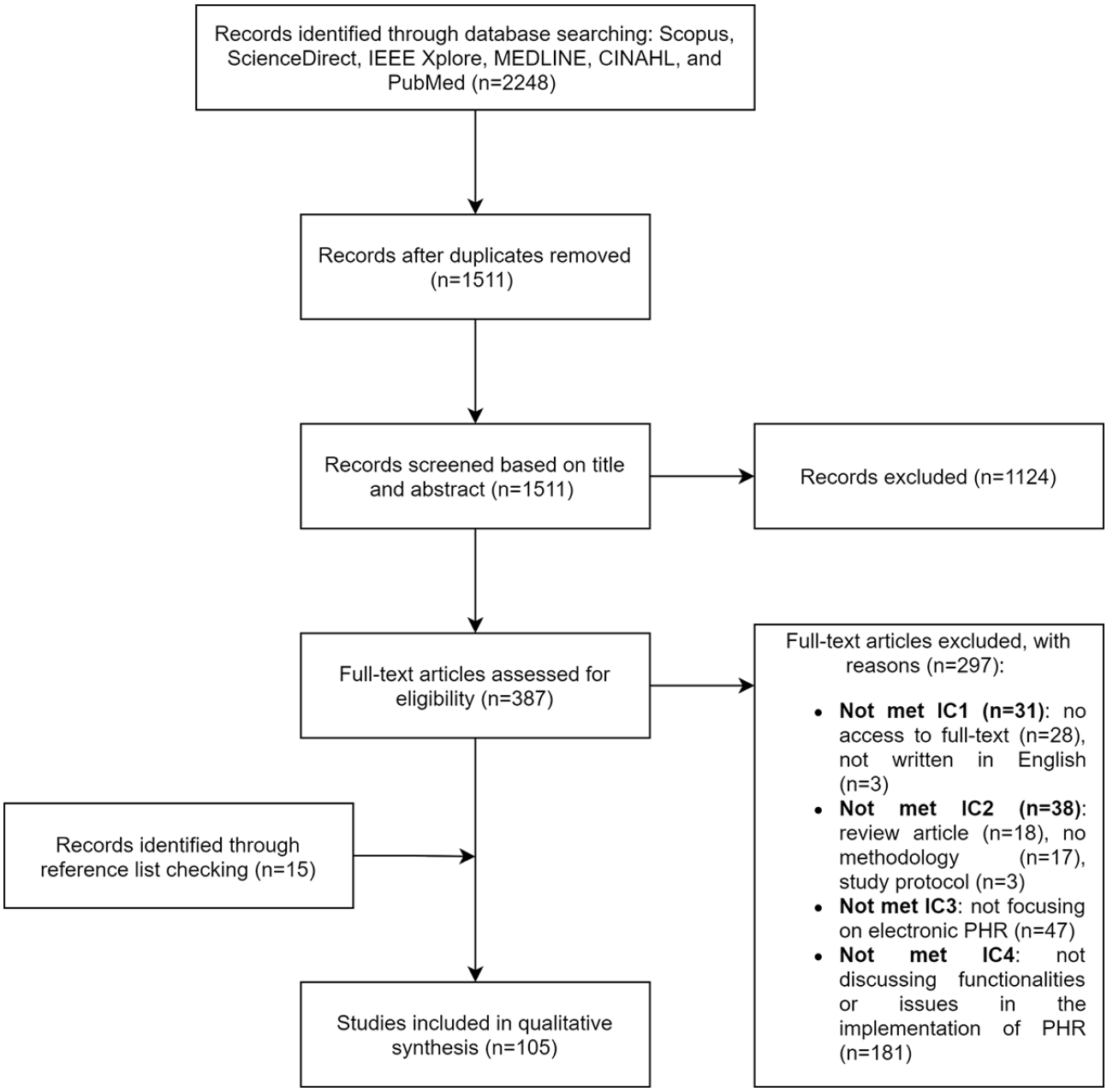
XSL•FO
RenderX

Results

Study Selection

The database search results identified 2248 studies from 2015 to 2020. Next, duplicate records were removed, resulting in a total of 1511 studies; 124 studies were excluded after the title and abstract screening (articles that mentioned literature review and articles not related to PHRs, patient portals, or access to health records to patients were excluded at this stage). A total of 387 articles were assessed in full text, of which 297 were excluded because they did not meet the selection criteria. However, 15 additional studies were identified from reference lists checking, for a total of 105 studies included in this review (Figure 1).

Figure 1. Flow diagram for search results. IEEE: Institute of Electrical and Electronics Engineers; MEDLINE: Medical Literature Analysis and Retrieval System Online; CINAHL: Cumulative Index to Nursing and Allied Health Literature; PHR: personal health record.



Study Characteristics

The chosen articles showed that PHR research has mainly been done in developed countries such as the United States, Canada, and European countries compared to developing countries. This country classification was based on the United Nations World Economic Situation and Prospects 2020 [21]. Countries involved in selected studies consist of developed countries such as the United States (42 studies), Canada (10 studies), Germany (8 studies), Australia (5 studies), Italy (4 studies), Netherlands (4 studies), United Kingdom (4 studies), South Korea (3 studies), European Union (2 studies), New Zealand (2 studies), Austria (1 study), Belgium (1 study), Norway (1 study), Portugal (1 study), and Taiwan (1 study) and developing countries such as Argentina (3 studies), China (3 studies), Iran (2 studies), Sri Lanka (2 studies), Brazil (1 study), Colombia (1 study), India (1 study), Malaysia (1 study), Romania (1 study), and Thailand (1 study; Figure 2).

XSL•FO
RenderX

The purposes of PHRs (Table 1) in selected articles include general, not specific to the disease, health status, or population (48 studies); chronic diseases such as cancer, cardiovascular disease, and diabetes (31 studies); hospital patients such as inpatients and outpatients (10 studies), older adults (5 studies), women and child health (4 studies), mental health (4 studies), and other specific populations such as employees and foster youth (3 studies).

The study methods (Table 2) used in selected studies include qualitative (41 studies), quantitative (33 studies), conceptual paper (16 studies), and mixed method (15 studies). A summary table of the characteristics of the included studies is provided in Multimedia Appendix 2.

Figure 2. Countries involved in personal health record study.

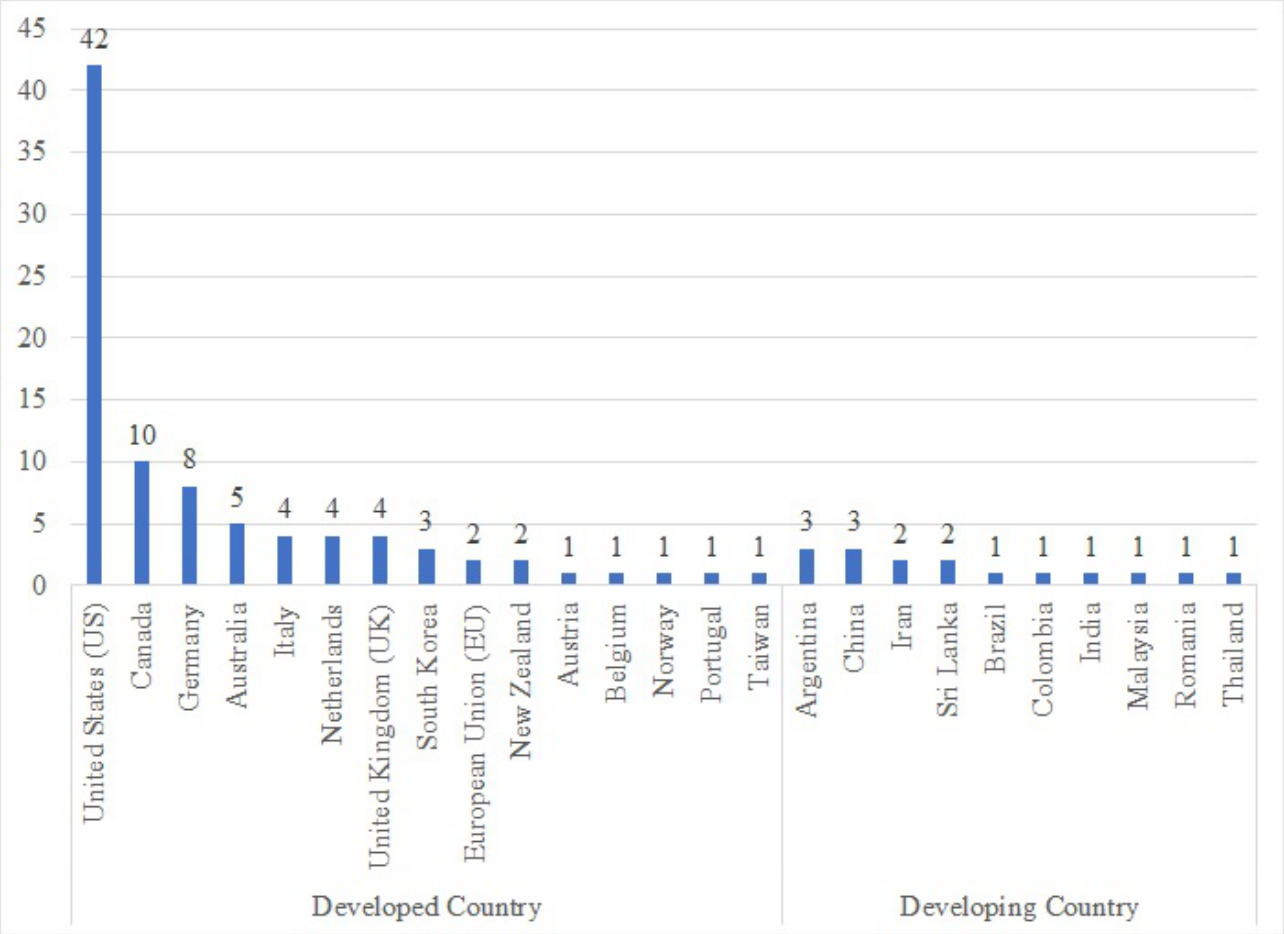


Table 1. Purposes of PHRs[a].

Category	Description	Number of studies
General	PHR designated not specific to any diseases, health status, or population.	48
Chronic disease	PHR for chronic diseases such as cancer, diabetes, or cardiovascular disease.	31
Hospital patients	PHR for patients who have visited the hospital, such as inpatients and outpatients.	10
Older adults	PHR for patients with the age of more than 50 years.	5
Women and child health	PHR for women, pregnancy, and pediatric health.	4
Mental health	PHR for mental health diseases, such as bipolar disorder.	4
Other populations	PHR for other specific populations, such as employees and foster youth.	3
Total	—[b]	105

[a]PHRs:patient health records.

[b]Not applicable.

XSL•FO
RenderX

Table 2. Methods used in the studies.

Method	Type of study	Number of studies
Qualitative	Interview and focus group discussion	41
Quantitative	Questionnaire, cohort study, and randomized clinical trial	33
Conceptual paper	—[a]	16
Mixed method	—	15
Total	—	105

[a]Not applicable.

Main Functionalities of PHR

Basic functions identified in this study consist of the health record and administrative record. Advanced functions consist of medication management, communication, appointment management, education, and self-health monitoring (Table 3). A summary table of the data elements and subfunctions is provided in Multimedia Appendix 3.

Table 3. Identified Functionalities in PHR[a].

Function	Description	References
Basic function		
Health record	Allows patients to view or access clinical documents from health providers' EHR[b].	[22-71]
Administrative record	Allows patients to manage personal information and view information related to health providers and insurance.	[22,25,26,30,31,34,39,42,44-46,55,59,66,68,70,72-78]
Advanced function		
Medications management	Allows patients to manage information related to medications and prescriptions.	[24-36,38-40,42,45,46,48,51-57,59-61,63,68-71,73,74,77,79-85]
Communication	Allows patients to interact and communicate with health care providers and others, such as support groups and families.	[22-24,27-32,36,38,40,42,43,46,48,49,51-53,55-57,59,61,62,67,69,71,74,75,80,82-84,86-94]
Appointment management	Allows patients to manage appointments with health care providers.	[22,23,25-31,33,34,36,40,42,47-53,55-57,59-61,63,64,67,71,73,77,78,81,85-87,92,94]
Education	Allows patients to access health-related education resources.	[22,30,31,40,45,46,55,57,59,61,70,71,76,77,90,95-99]
Self-health monitoring	Allows patients to manage their self-health data through clinical measures.	[23,26,30,33,39,44,58,66,67,70,72,81,85,86,90,93,95,96,98,100-105]

[a]PHR: patient health record.

[b]EHR: electronic health record.

Basic Functions

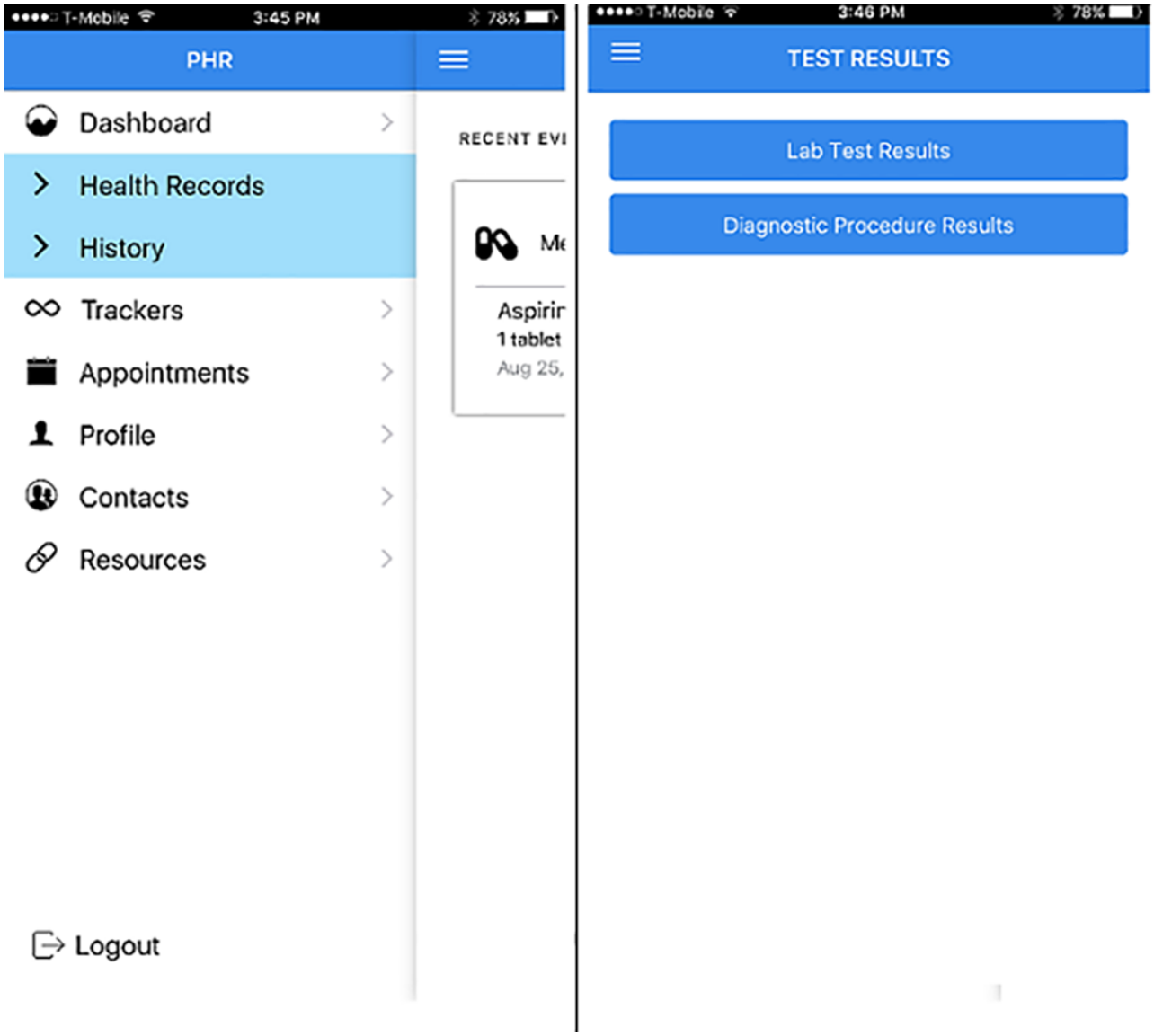
Health Record

The health record function provides patients options to view clinical documents that can be retrieved from health providers' EHR [22,23,34]. This information can include problem lists [22,24-29,45,56,67-71], allergies [22,25-28,30-33,35-39,56,67-70], immunization [22,26-28,30,32-35, 40-42,56,68-71], laboratory and test results [22,24,26-30,33,36,37,40,42-61,67-71], diagnostic information [32,37,44,45,62], discharge information [31,34,63], and clinical notes [24,30,33,42,50,61,64,67,69]. Figure 3 shows an example of test results in PittPHR [33].

This function can also include information about medical history [22,24,30,33,35,37,39,45,51,57,64,65,67,68], family history [28,30,32,33,35,37,66], genetic history [45], surgical history [26,28,33,35,45,66,68], social history [32,33,35,37,45,68]. Some studies suggested that this function also supports patients' ability to print the record [24,36] and add comments or notes in health records [28,61].

XSL•FO
RenderX

Figure 3. Test results menu in the health records module of PittPHR [33].



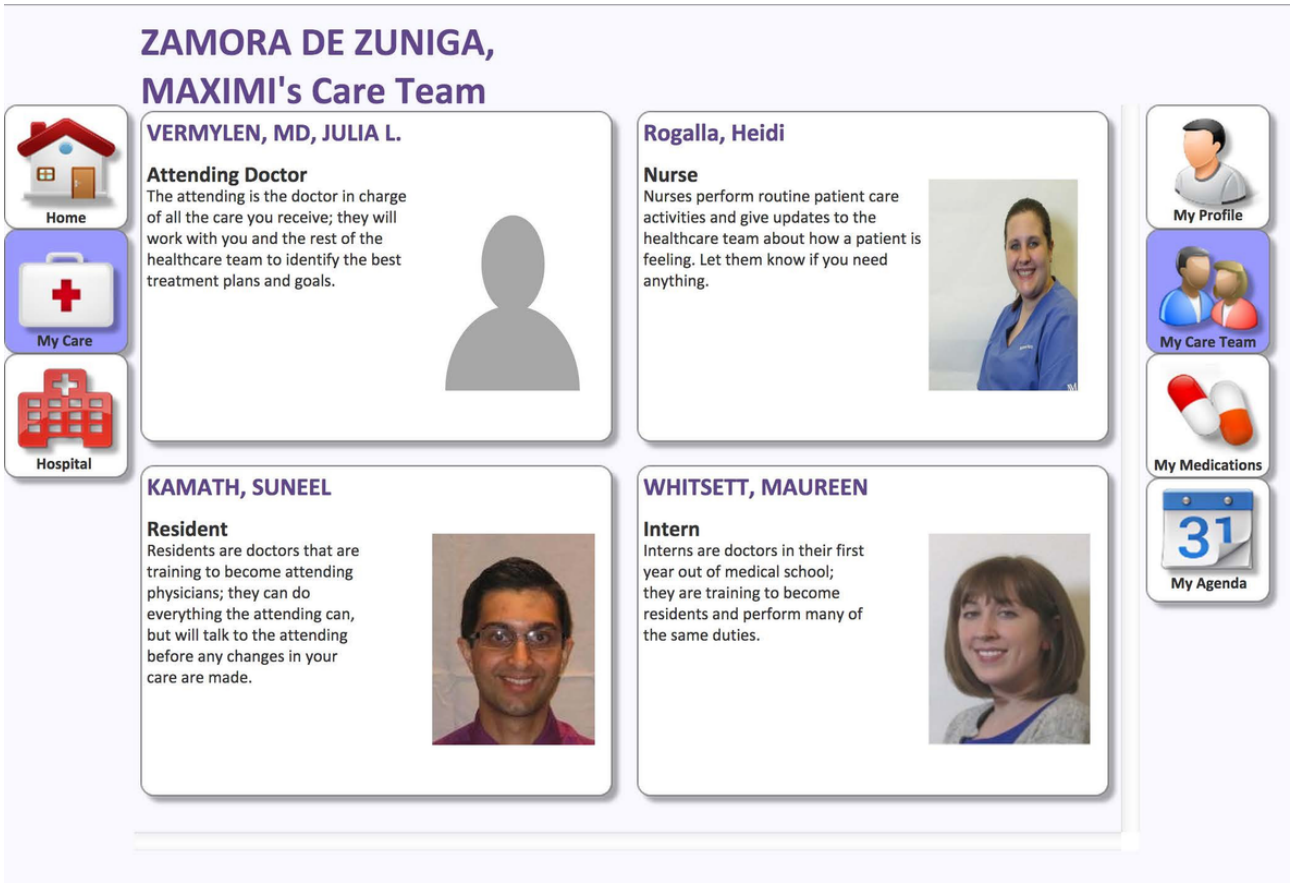
Administrative Record

This function enables the patient to manage information about demographics or personal information [25,26,31,44,45,68,72] such as name [25,39,44,72,73], gender [25,26,44,72], birthdate [25,26,44,72,73], blood type [39,44], contact information [25,31,68,72], and parents' names [25,45]. Patients can also change their information, such as change password, address, and email address [22]. Patients can enter this information [22,31,68,72] or retrieve it from a central patient registry, such as in Lifelong PHR [34].

Patients can also view health professionals' information, such as name of health worker [34,39,55,74], role [74,75], educational background [75], contact information [42,45], specialty [76], location [76], and pictures [46,55,59,74,77,78]. Patients can also view hospital information, such as location, contact info, address, navigation [70]. Patients can also view and pay bill [42,55,77] and get insurance-related information [22,30,34,45,66]. This data can be retrieved from the regional health care information system [34]. Figure 4 shows an example of the health care team information in the PHR app [74].

XSL•FO
RenderX

Figure 4. Health care team information menu [74].



Advanced Functions

Medication Management

Health care providers publish prescriptions to the patient's PHR, while pharmacists dispense the prescribed medication [34]. This function provides information about the list of medications that patients are currently taking [25,28,29,35,38,45,46,52,56,59,60,68,69], medication name and dosage [32,35,40,60,74,77,79,80], and list of past medications [28,29,36,42,45,46,60,69,70,73,81]. PHRs should also add information about the purpose or class of medications to give patients an understanding of the medication type [74] and allow pharmacists to explore the data according to their common questions [80].

This function also allows patients to view list of prescribed medications [26,30,39,54,79,82,83], prescribing physician [79], refill prescription [24,27,29-31,33,34,36,39,48,53,55,57,60,61,84], order medications [29,39,71], deliver purchased medication [79], as well as track the delivery of medication [36,39,42]. Some PHRs also provide medication schedulers and reminders of when to take medicines [28,70,83,85], drug or medicine reconciliation [42,51,63,83], and warning alerts of potential adverse interactions based on the medication and allergy list [38,68,73]. Figure 5 shows an example of medication management in medication management in My Chart in My Hand [85].

XSL•FO
RenderX

Figure 5. Medication management in My Chart in My Hand [85].

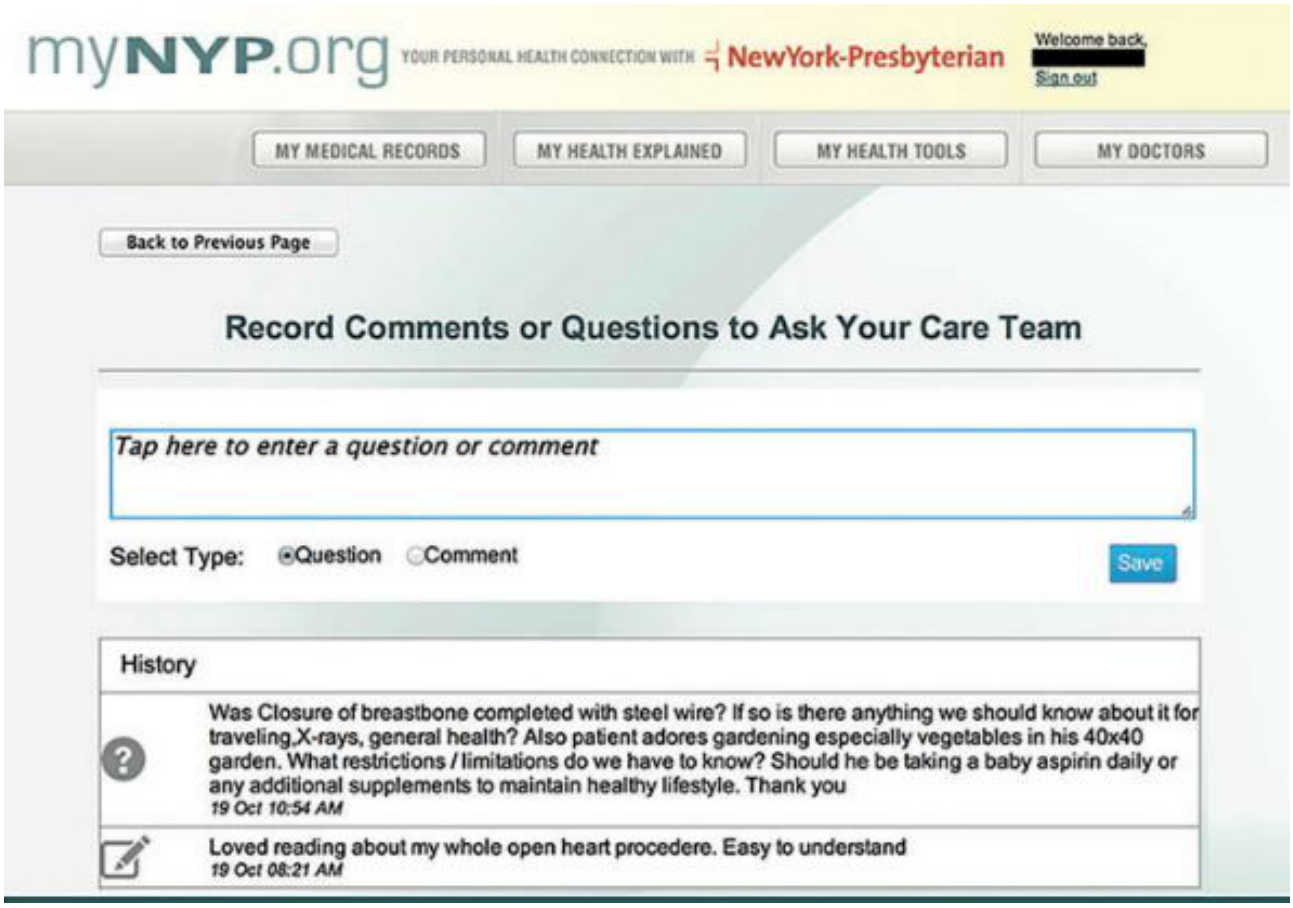


Communication

The patient can send messages to the health care provider to inform them of health condition [23], share doubts and worries [86], receive medical advice [56,86], or send nonurgent messages [40,46,71]. The communication can be in the form of messaging [23,27,29-32,36,38,40,42,43,46,48,49,51-53,55-57,59,61,67,69,71,74,75,82-84,86-92] or text where patients can write questions (Figure 6) [74,80]. Some PHRs also enable patients to contact others in a similar situation [28,29,83,88], support groups [62,87], family [75,89], or customer support and billing departments [22]. Some studies also suggested this function have the ability to maintain a record of past conversations [36] and provide email or text notification when a health care provider leaves a message on the PHR [24,93,94]. Moreover, some studies suggested tracking the status of a question [80], message multiple providers at the same time [24], and import selected emails and interactions on the social network to PHRs [86].

XSL•FO
RenderX

Figure 6. Comments or questions page in myNYP.org [80].



Appointment Management

Some PHRs may allow a patient to request or schedule appointments (Figure 7) [22,23,25-28,30,33,34,40,42,48-50,52,55,57,60,61,67,71,85,87,92], while others only allow patients to view their past and upcoming appointments [29,31,36,51,53,56,59,63,64,73,77]. The types of appointments can include patient-doctor visit consultation services and other health services such as specialist encounters, sample takings, hospital admissions, result withdrawal [86], therapies, and online consultation [23]. Moreover, some studies suggested that PHRs include reminders or notifications for upcoming appointments [33,42,47,48,60,61,81,94]. This reminder can be in the form of email notifications about the date and time of the appointment [42,81]. PHRs can also add a calendar to keep track of future appointments [34,78].

XSL•FO
RenderX

Figure 7. Appointment scheduling in mPHR [25].

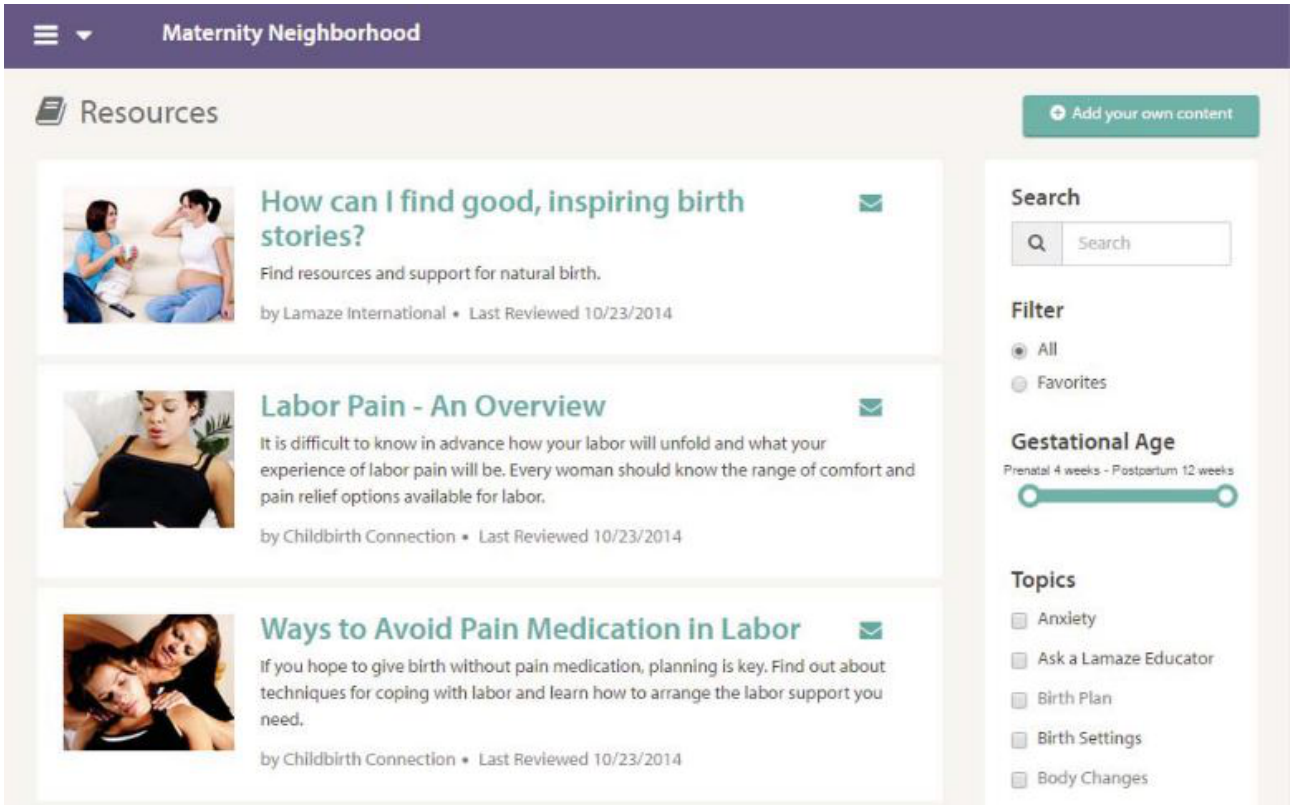


Education

The education function can include resources from trusted websites [45,90], health information libraries [22,30], video resources [46,59,95], or government supported information [95]. The information can consist of lifestyle management [45,57,71], first-aid information [40,70], discharge instructions [31], surgical procedure [77], physical activities guidance [96], or health-specific education such as pregnancy [97,98], mental health [45,61], or chronic diseases–related education [90,95]. Figure 8 shows an example of the education page in the Maternity Information Access Point [97]. Health providers are responsible for providing clinical topics and resources for credible information [55,76,99]. Moreover, PHRs should also have the ability to search for information using an intelligent search engine [99].

XSL•FO
RenderX

Figure 8. Education resources in Maternity Information Access Point [97].



Self-Health Monitoring

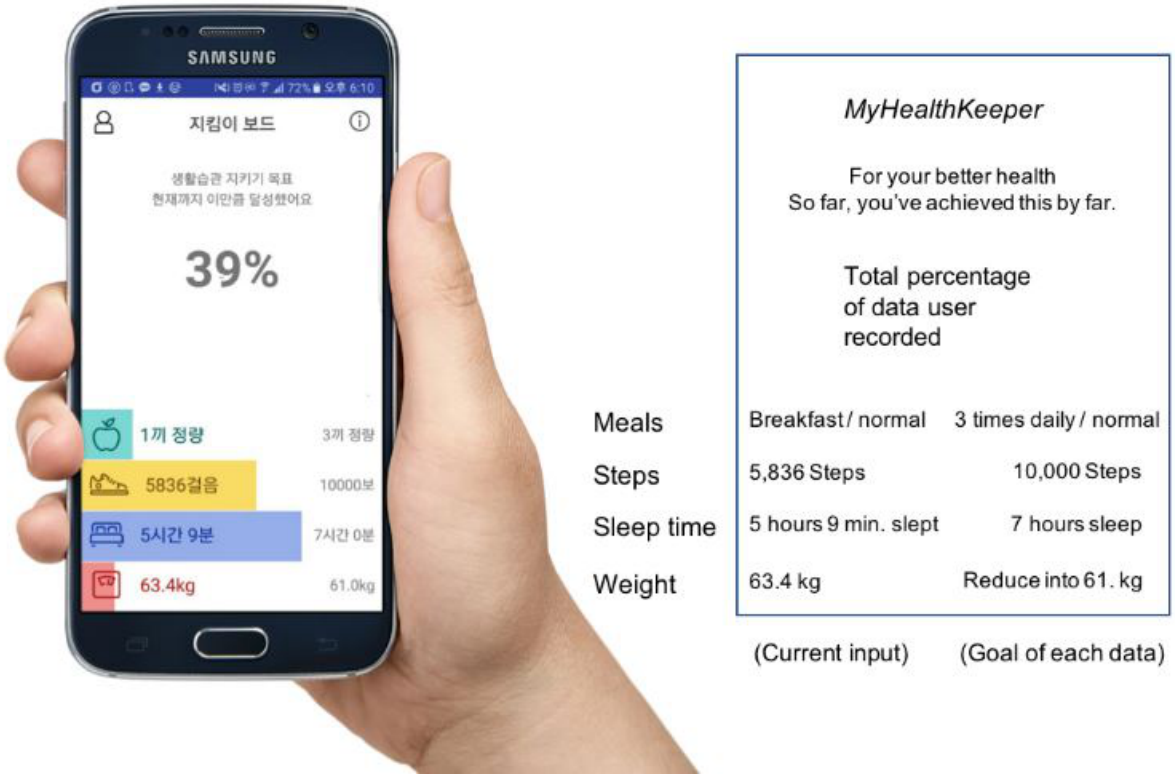
Patients can manage their own health related to nutrition and diet information such as weight [30,33,39,58,66,67,72,85,86,90,98,100-102], height [39,58,66], physical activity or exercise [30,33,58,66,70,96,98,100,101], and food and meals [33,66,98]. Patients can also manage their vital sign data such as temperature [26,44], blood pressure [30,33,44,58,66,67,70,72,85,90,98,101,103], blood glucose [30,58,66,70,72,85,86,98,103], and heart rate [90]. Patients can also monitor other self-health data such as sleep [33,66,95,100,101], period [33,100], moods [98,100,101], and stress [66,70,100].

These clinical measures enable calculation such as BMI [39,66,70,85,100], body fat percentage [70], waist-to-height ratio [70], calorie [70], cholesterol level [66,86], and glycemia [86]. This information can also calculate disease risks such as cardiovascular disease risk and metabolic syndrome risk [85]. The data in this function can be retrieved from home monitoring devices [23,39,58,85,96,104,105] such as Bluetooth-enabled health monitors [104], accelerometers [105], blood pressure monitors [58], blood glucose meters [58,81,85,93], and pedometer [72] and fitness tracker apps [39].

The monitoring of health data can be shown as a dashboard that visualizes data in graphs, charts, or diagrams [30,33,70,72,81,86,101,102,105]. Key performance indicators can be different for each patient, depending on their conditions. For example, in the MyHealthKeeper app, the clinician provided individual diet and physical activity targets for each patient during an outpatient visit (Figure 9) [101]. This function can be integrated into a clinician's EHR, and clinicians could review these data and provide feedback about the health-related lifestyle management of their patients [101].

XSL•FO
RenderX

Figure 9. MyHealthKeeper interface for patient's lifestyle data [101].



Issues in Implementation of the PHR

Some issues must be considered in implementing PHRs because these issues can define additional functionalities that can support the main functionalities in PHRs. The issues identified included interoperability, security and privacy, usability, data quality, and personalization (Table 4).

Table 4. Issues in implementation of the PHR[a].

Issues	Description	References
Interoperability	Ability of PHR to share or exchange data with other systems	[22,25,29,33,35,37-39,51,53,61-63,68,72,76,86,88-90,96,100,102,104,106-112]
Security and privacy	Safeguarding of data and personal information in PHR	[25,26,32-34,42,45,47,54,55,57,60-65,69,70,72-74,76,77,83,92,94,97,103,108,110,112-122]
Usability	Whether users can use PHR effectively and efficiently	[24,25,33-37,41,45,46,48,51,54,59,61,62,65,71,74,76,81,83,85,87,90,92-94,96,100,105,108,110,117-126]
Data quality	Ensures consistency, completeness, accuracy, and timeliness of the PHR information	[24,30,31,33-35,64,68,81,90,91,107,108,110-113,115]
Personalization	Ability of PHR to be tailored and adapted to patient needs and preferences	[22,33,52,57,59,78,87,88,95,99,102,112,117,118,123,125]

[a]PHR: personal health record.

Interoperability

An important issue raised in several studies is PHR compatibility with other systems [51,88,90,100,106]. Health service providers such as clinicians should input data from other systems into the PHR or vice versa, which was considered too time-consuming and unfeasible for daily practice [100]. This problem can also be caused by health organizations adapting their formats to use health records and not allowing health information sharing in their PHR to other applications or organizations [63,107]. As a result, a patient may have health records scattered in several applications [107]. To provide more benefits and ensure its successful implementation, PHRs should realize interoperability among various data and systems [106,108]. PHRs should have the ability to share information with others [88], such as health professionals [37,61,89].

In the tethered or integrated PHR, patients may connect their PHR to the health care provider system [25,62,72,76,86,102,104]. With this integration, health information is automatically transferred to the PHR [33,39,62,109,110]. This can reduce data entry load [33,96], improve data accuracy [62,96], prevent medical errors [38], reduce the health information recall [35], and contribute to users' better perceptions about the system's usefulness [111]. It is also suggested that PHRs be integrated into various health providers

XSL•FO
RenderX

and not limited to one health provider [53,102]. Patients may also have the ability to share information with trusted institutions and insurance bodies to speed up reimbursement procedures [86] and access other family members' records [22,102].

It is necessary to create legislation to realize PHR interoperability [106]. Health providers need to provide standard definitions for data exchange and cooperate with other providers [63]. There are international standards or frameworks for interoperability, such as OpenEHR, Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR), and Integrating the Healthcare Enterprise and Continua Health Alliance specifications [29,68,72,107,112]. OpenEHR describes the management and exchange of data in EHRs for developing PHRs using specific language [68]. OpenEHR integrated with other standards in particular health data types, such as laboratory results [107]. Similarly, HL7 FHIR enables the management of a single data entity, group of entities, or a record using well-known standard languages [68]. FHIR application program interface allows any arbitrary system connected with another medical system already equipped with the FHIR application program interface [72]. FHIR allows the patient portal to be interconnected but independent [29]. Moreover, Integrating the Healthcare Enterprise specifies architectural approaches using international standards for the health data exchange and can fit the mobile platform's resources. At the same time, Continua enables communication from personal health devices to EHRs and PHRs [112].

Security and Privacy

PHRs contain personal and sensitive data [47,77,108,112-114]. Some people have concerns about storing these data online [54,103,108,113,115] and consent to use the system [116]. They may have concerns about identity theft and unauthorized access in PHRs [54,55,57,61,69]. Confidentiality and privacy of information in PHRs should be ensured through secured access to PHRs [110].

To ensure the security of information, PHRs should use a single sign-on mechanism [70], user authentication [26,33,64,72,73,112,117], authorization [42,112], identity verification [34,63], encryption [25,33,112,118] or pseudonymization [114], backup mechanism [25,33,72], and firewalls [72]. PHRs can also implement an access log so that users can see who viewed and downloaded information [76]. The use of complicated or complex passwords can improve the security of s [47,119]. However, some studies show that users have difficulty remembering their passwords [47,74,77,92,94,97,103,120]. Thus, PHRs should also add other methods such as fingerprint authentication [97], biometric identification [33,94], citizen digital certificate [121], and allow users to change their passwords [62,72].

To address privacy concerns related to data sharing, PHRs should have the ability for patients to choose what information to share and who can see that information [34,42,45,60,61,65,76,83,121,122] and provide a privacy policy in the system [32]. The consent model should also be considered in implementing PHRs [116]. Moreover, PHR systems need to follow specific legal requirements related to security and privacy defined on regional, national, or international levels [112]. For example, the Health Insurance Portability and Accountability Act ensures secure data exchange with entire clinics [76].

Usability

Some usability problems identified in selected studies include font or text size that are difficult to use [71,94], confusing format [81], unclear visualization of data [90], problem with navigation [51,59], and complicated data entry [85,118]. Complicated data entry may cause users to not enter data correctly into their PHR [118]. The reduction and simplification of PHR system data entry should be considered in PHR design [35,93,118]. Users prefer easy to use, simple, and user-friendly interface [24,41,45,54,61,62,65,92,94,110,118,120-124]. Users are also interested in attractive and interactive systems [25,33,108,110,120,124] such as the use of contrasting colors for scroll bars and menu items [59]. Moreover, it is also important to maintain consistency and standardization of interfaces [35,74,117,118]. A mobile app version of the PHR was also suggested because it was perceived as more user-friendly and easy to use [25,34,61,65,81,93,100].

A PHR may add a section to guide patients about the features in the PHR [46,54,59,61,65,83,90,96,118] and quick access to the essential functionalities [37,48,108]. The use of user-interface elements like buttons and a dropdown menu can enhance the user-friendliness and simplicity of the PHR interface [25,100]. However, icons should be avoided when designing for older adults since they may not recognize them [96]. PHRs should be easy to understand and navigate for all user groups [110], including those with basic computer knowledge and those who are not computer literate [94,118]. PHR usability should be determined using health literacy assessments and there should be different PHR versions for specific groups of users [36].

Developers should involve users in designing, updating, or improving PHR systems [48,71,119]. Using a user-centered design approach can facilitate users' involvement in PHR design [76,87]. The user-centered design process increased the development process's complexity, but the product quality was higher, especially satisfaction and user acceptance [105]. However, user-centered design may not apply to all PHR types, especially PHRs targeting the general population, which necessitates identifying specific user groups and specific use contexts [125]. Adopting a usability design framework that includes usability and user testing may help address PHR usability issues [126]. Standardization used for PHR design is International Standards Organization (ISO 9241-210), which focuses on the requirements and user needs [105], and ISO 9241-11 for software systems components that define usability [96].

Data Quality

Health care providers may doubt patient-entered data in a PHR [30,113]. Not all patients have enough knowledge to generate health data in a PHR [107,112]. Data uploaded by the patient may be inconsistent [91], incomplete [81,90,110], inaccurate [30,81,90,110,115], or not up to date [90]. PHRs require patient commitment to keep the system up-to-date and relevant over time [111]. This issue needs particular attention, especially

XSL•FO
RenderX

when PHR data are transferred to EHRs and used in professional medical decision-making and treatment processes [112].

To ensure data quality of patient-generated data in PHRs, health care professionals need to take time to supervise the quality of information generated by patients in PHRs [110,112]. PHRs should differentiate patient-generated data from the health care provider's data [68]. Moreover, PHR design needs to define what information is required because an incomplete record is preferable to an inaccurate one from a provider's perspective [35]. Standardization of patient-entered information is essential to ensure data quality [33]. Input control should be comfortable and descriptive words should appear to help patients enter PHR data [108].

In tethered PHRs, which are tied to EHRs in health organizations, health information on the PHRs are created automatically from the original patient clinical reports to make this information more reliable [34]. However, this can be a problem if the EHR's information is incomplete [30,64] or if the information is not generated automatically. This can also be caused by health care providers not updating the PHR information consistently [24,31].

Personalization

Some users may have more health issues than others, such as older patients having more health issues, appointments, and information to manage [22]. People want the PHR to be tailored to their needs and capable of changing based on their health and well-being needs [59,78,88,102,123]. This person-specific health and well-being information can make the PHR system more appealing [118]. The PHR system needs to be adaptable and extensible to ensure successful operation [112]. It is also suggested that PHRs support customizability based on computer literacy [87].

PHR systems should provide medical information that can be dynamically adapted to patient preferences for simpler or more complex information [99] [117]. For example, in PittPHR, users can customize the trackers according to their own needs by hiding or unhiding available trackers in a given list and add or delete links in the resources module according to their own needs [33]. PHRs could also provide tailored health education materials based on patient health problems [52,57,95,117]. Despite the need for personalization, designers or developers need to define the extent to which PHRs can be personalized but still maintain standardization, uniformity, and simplicity [125].

Discussion

Principal Findings

Seven function categories of PHRs are identified as the main functionalities of PHRs, which are grouped into basic and advanced functions. Basic functions (health records and administrative records) provide essential information for patients in their health care. Health records could provide a complete summary of patient health status and condition. Information on this function could reduce health workers' time gathering patient history and reduce redundant transactions and tests [20]. Information on administrative records such as personal information serves as a patient identifier on a PHR.

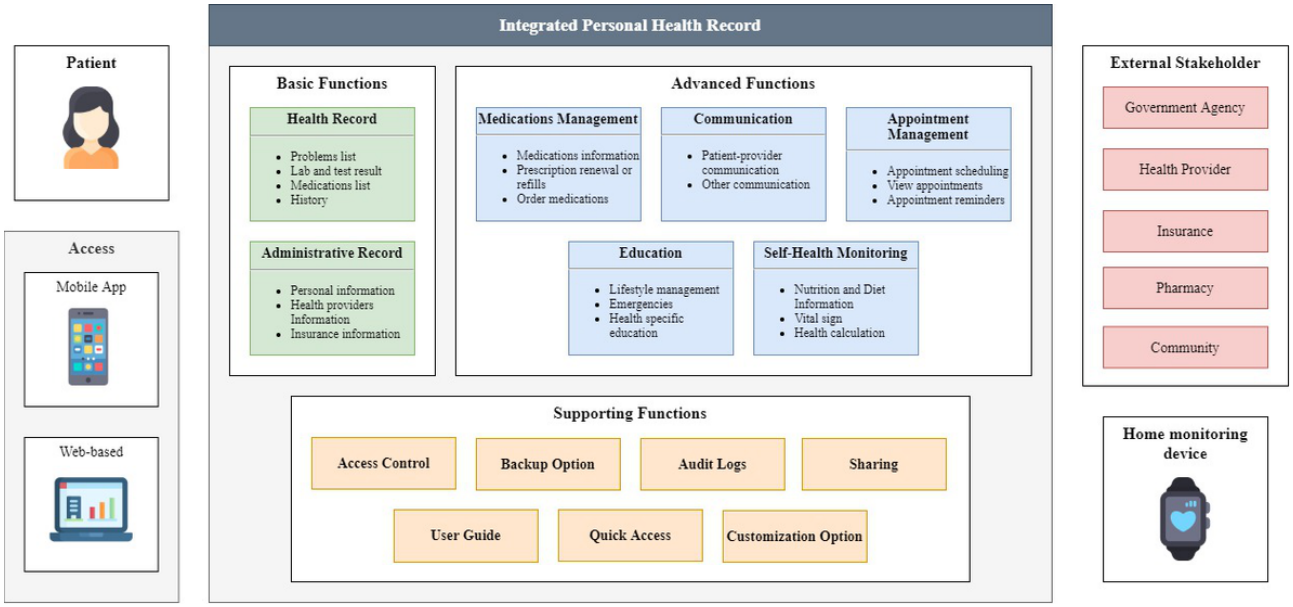
Advanced functions (medication management, communication, appointment management, education, and self-health monitoring) could support patient involvement in their health care. Involving patients in controlling their health information improves the chance that health providers would have a comprehensive view of patient health conditions [20]. Medication management functions such as medication scheduler and reminders could help patients take medicine on time. Moreover, the medication reconciliation option could avoid medication errors [127]. Communication functions such as messaging could free physicians from the limitations of phone and face-to-face communication [1]. Appointment management reduces the chance of a patient missing an appointment. Education could support health knowledge promotion [20], which may improve the patient's health literacy. The information recorded from the self-health monitoring function may help health providers with disease diagnosis and treatment [10]. This function could help patients track their progress to reach specific health goals [9] and monitor the impact of their behavioral changes [12].

In addition to these functionalities, other supporting functionalities may also need to be developed based on the issues identified in this study. To improve security and privacy, PHRs should implement access control, which includes authentication and authorization. PHRs can also provide a backup option to avoid data loss and audit logs to review who accessed the record and what data have been accessed. To improve usability, PHRs can provide quick access to the important information or functions that users frequently use and add a menu for help or a user guide about using features in the PHR. Customization options to show or hide specific health data according to patient health needs are also recommended to increase personalization.

Interoperability represents a key component of PHR architecture [8]. When PHRs are integrated with health providers, they provide more significant benefits and valuable content for users [1,20]. Interoperability can also reduce data entry load because health information is automatically transferred to the PHR. This can increase the usability and the quality of data on the PHR. PHRs need to provide standard definitions for data exchange and implement sharing functions to connect PHRs with other stakeholders, such as health providers, insurance, government agency, pharmacy, community or support groups, and other systems such as home monitoring devices. Figure 10 describes the integrated PHR architecture based on the result of this review.

XSL•FO
RenderX

Figure 10. Integrated personal health record architecture.



The main functionalities described in this review, such as health records, administrative records, medication management, communication, appointment management, education, and self-health monitoring, have also been described in previous reviews [9-12]. Most of these reviews [9,11,12] focus more on discussing the functionality of PHRs related to improving health service delivery. Only Genitsaridi et al [10] discussed supporting functionalities such as access control to be included in requirements on the PHR. Previous studies also have discussed technological issues [13-15]. Our research augments that of previous studies by translating these issues as supporting functionalities in PHR systems.

The functionalities in PHRs can help health care providers and patients obtain useful health information during public health emergencies such as natural disasters and pandemics. For example, in the COVID-19 pandemic, hospital services experienced a crisis [128]. Observations of health outside of standard hospital settings can be difficult [129]. Functions such as communication can help patients and health care providers consult without making eye contact. Furthermore, in the education function, PHR providers can provide information about updated COVID-19 and health care information. The health record function helps patients obtain and store test results. The self-health monitoring function increases the patient's ability to control and manage health conditions. Functions such as measuring body temperature can be used to detect early signs of infection [129]. Integrating PHRs into a broader telehealth infrastructure could improve emergency health care delivery by reducing patient spikes in health care facilities [2].

Comparison With Prior Work

The Health Level 7 Personal Health Record System Functional Model (HL7 PHR-S FM) defines a standardized model of the functions present in PHR systems [130,131]. The model consists of 3 sections: personal health (PH), supportive (S), and information infrastructure (IN). Personal health functions enable an individual to manage information about their health care. Supportive roles assist with the administrative and financial requirements within health care delivery. Information infrastructure functions support personal health and supportive functions.

Health records, medication management, communication, education, and self-health monitoring can be categorized into personal health sections. Administrative records such as managing patient profiles can be categorized into a personal health section, while information about health professionals, hospitals, and insurance can be categorized as a supportive section. Supporting functions defined based on PHR implementation issues, namely sharing, access control, audit logs, backup options, and customization, can be categorized in the information infrastructure section. This section ensures the privacy and security of PHRs, promotes interoperability between PHRs and other systems, and enables PHR function to be accessible and easy to use [130,131]. Table 5 summarized comparisons between functions identified in this review study and functions defined in the HL7 PHR-S FM.

XSL•FO
RenderX

Table 5. Comparisons between functions.

Functions identified and ID	Functions defined in the HL7 PHR-S FM[a]	
	Function name	Description
Basic functions		
Health record		
PH.2.5	Manage historical and current state data	Provide a summary of the patient's current medical state and history
Administrative record		
PH.1.2	Manage PHR[b] account holder demographics	Capture the patient's demographic information
S.1.3	Manage health care provider information	Import or retrieval of data necessary to identify a health care provider
S.1.5	Manage health care facility information	Import or retrieve of data necessary to identify a health care facility
S.2.1	Capture and read health insurance account and benefit information	Request and/or receive and read the information on health insurance benefits
Advanced functions		
Medications management		
PH.3.4	Manage medications	Help patients manage his or her medications
Communication (patient-provider communication)		
PH.6.3	Communications between provider and/or the PHR account holder's representative	Capture information in preparation for a consultation and maintain continuous communications with the health provider
IN.3.10	Secure messaging	Enable secure electronic communication with health providers
Appointment management		
PH.6.3	Communications between provider and/or the PHR account holder's representative	Capture information in preparation for a consultation and maintain continuous communications with the health provider
Education		
PH.4	Manage health education	Provide proper medical education and patient-specific knowledge based on information in the PHR
Self-health monitoring		
PH.3.1	Manage personal clinical measurements and observations	Provide the patient capability to enter personally sourced data and make it available to authorized health providers or other users or applications
Supporting functions		
Sharing		
IN.2	Standards-based interoperability	Interoperability standards enable the sharing of information between PHRs and other systems
Access control		
IN.3.3	Entity access control	PHR must perform authentication and authorization of users or applications
Audit logs and backup option		
IN.4	Auditable records	Provide system access and use audit capabilities to indicate who accessed the record, how, and when the action was taken
Customization option		
IN.1.3	Present ad hoc views of the health record	Provide ad hoc views of the PHR information
User guide		
PH.1.1	Identify and maintain a PHR account holder record	Offer user guide for the installation, initialization, registration, or operation of their PHR

[a]HL7 PHR-S FM: Health Level 7 Personal Health Record System Functional Model.

[b]PHR: personal health record.

XSL•FO
RenderX

The functionality identified in this review covers the main section (PH, S, IN) in the HL7 PHR-S FM. However, functionalities and data elements found in this review are on the individual level that focuses on improving health care. Functions that are not included in this review are functions related to the secondary use of health data. Secondary health data use applies to personal health information for uses outside direct health care delivery [132]. In the HL7 PHR-S FM, a population health and wellness (PH 3.6) function helps control public health risks to the population and patients. For example, it enables patients to export anonymized data for biosurveillance and public health reporting, and patients can get alerts or warnings regarding population health threats. A manage other resources (S.4) function supports patient enrollment in clinical trials or research [131]. From this review, only a few studies [34,91] mentioned that PHRs could be used for secondary health data use, but they did not explain specific data needed for this function. A discussion about secondary health data use in PHRs can be an opportunity for future research.

Not all functions in the HL7 PHR-S FM were found in this review study because the HL7 PHR-S FM is universal and generic by design. There may be additional constraints in certain realms or regions. PHR developers or designers can create a functional profile to define a selected set of applicable functions for a particular purpose, group of users, degree of interoperability, or custodian [130]. This study defines PHR functionalities based on the current state of research and provides more examples of data elements and subfunctions for each functionality. This study also found that the HL7 PHR-S FM only includes patient-provider communication. Other communications, such as communication with others in a similar situation and support groups, are not discussed in the HL7 PHR-S FM.

Limitations

This study is limited to reviewing the implementation of PHRs in research articles and does not address the implementation of commercial PHRs available on the internet. Thus, the functionalities and issues of the PHRs defined in this study may not reflect the state of the practice. This paper does not discuss which functions are more common or whether certain functions are used more frequently than others and does not discuss each function's benefits and impact on health outcomes. We cannot determine which functionality should be prioritized in the implementation of PHR. We only discuss the functions that are generally mentioned in the selected paper. Each function's data element may not be comprehensive and might not be generalizable to all patient populations. This is because each disease or condition has different specific data.

Conclusions

This systematic literature review paper discussed functionalities and issues in the implementation of PHRs. Seven function categories are identified in this review, which are grouped into basic and advanced functions. In addition to these functionalities, other supporting functionalities may also need to be developed based on the issues identified in this study. Based on the results, this paper provides an integrated PHR architectural model that describes the functional requirements and data sources of PHRs. This study can offer recommendations or guidance in implementing PHRs by health care facilities management, application developers, policymakers, or other related stakeholders. Functionalities (including data elements and subfunctions) listed in this study and architectural model (Figure 10) can be used when considering what features to implement in a PHR. The model (Figure 10) can also serve as the target data sources to be integrated into the PHR system. Moreover, technological issues explained in this study can be used to develop policies in the implementation of PHRs. For example, since security and privacy are identified as technological issues in this study, implementers of PHRs should develop policies that govern access control in PHRs. The findings of this study may be translated as functional and nonfunctional requirements of the PHR system. This study's findings can also serve as a basis and comparison for other researchers who will examine PHR functionality and use in the future. PHR integrated architecture (Figure 10) can be used as a model that other researchers can use to compare, map, or evaluate the PHR functionalities that will be examined. Furthermore, personal factors such as age, culture, and health and technology literacy levels can influence security, privacy, and usability issues. Future studies can be conducted to analyze the effect of personal factors on technological issues.

Acknowledgments

This work is supported by grant NKB-3027/UN2.RST/HKP.05.00/2020 from the Pendidikan Magister Menuju Doktor untuk Sarjana Unggul and Ministry of Research, Technology, and Higher Education Republic of Indonesia.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOC File , 66 KB - jmir_v23i7e26236_app1.doc]

Multimedia Appendix 2
Characteristics of the included studies.

XSL•FO
RenderX

[DOCX File , 50 KB - jmir_v23i7e26236_app2.docx]

Multimedia Appendix 3
Data elements and subfunctions.
[DOCX File , 358 KB - jmir_v23i7e26236_app3.docx]

References

1. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
2. Bouri N, Ravi S. Going mobile: how mobile personal health records can improve health care during emergencies. JMIR Mhealth Uhealth 2014;2(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3017] [Medline: 25098942]
3. Perakslis E, Huang E. Covid-19: a huge stress test for electronic health record systems. Stat. 2020 Mar 12. URL: https://www.statnews.com/2020/03/12/covid-19-huge-stress-test-electronic-health-record-systems/ [accessed 2020-04-06]
4. Finnegan M. COVID-19 puts new demands on e-health record systems. Computerworld. 2020 Mar 31. URL: https://www.computerworld.com/article/3535083/covid-19-puts-new-demands-on-e-health-record-systems.html [accessed 2020-04-06]
5. World Health Organization. Management of patient information: Trends and challenges in Member States. In: WHO Global Observatory for eHealth. Geneva: World Health Organization; 2012.
6. Galas J. COVID-19 pandemic: why patients need to control their own data. HIT Consultant Media. 2020 Mar 24. URL: https://hitconsultant.net/2020/03/24/covid-19-pandemic-patients-need-control-their-own-data/#.YObQqugzbIU [accessed 2020-04-07]
7. Alsahafi A, Gay B. An overview of electronic personal health records. Heal Policy Technol 2018 Dec;7(4):427-432. [doi: 10.1016/j.hlpt.2018.10.004]
8. Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW. A research agenda for personal health records (PHRs). J Am Med Inform Assoc 2008;15(6):729-736 [FREE Full text] [doi: 10.1197/jamia.M2547] [Medline: 18756002]
9. Bouayad L, Ialynytchev A, Padmanabhan B. Patient health record systems scope and functionalities: literature review and future directions. J Med Internet Res 2017 Nov 15;19(11):e388 [FREE Full text] [doi: 10.2196/jmir.8073] [Medline: 29141839]
10. Genitsaridi I, Kondylakis H, Koumakis L, Marias K, Tsiknakis M. Evaluation of personal health record systems through the lenses of EC research projects. Comput Biol Med 2015 Apr;59:175-185. [doi: 10.1016/j.compbiomed.2013.11.004] [Medline: 24315661]
11. Alturkistani A, Qavi A, Anyanwu PE, Greenfield G, Greaves F, Costelloe C. Patient portal functionalities and patient outcomes among patients with diabetes: systematic review. J Med Internet Res 2020 Sep 22;22(9):e18976 [FREE Full text] [doi: 10.2196/18976] [Medline: 32960182]
12. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention: a systematic review. BMC Med Inform Decis Mak 2015;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
13. Hawthorne K, Richards L. Personal health records: a new type of electronic medical record. RMJ 2017 Nov 20;27(3):286-301. [doi: 10.1108/RMJ-08-2016-0020]
14. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
15. Carrión SI, Fernández-Alemán JL, Toval A. Are personal health records safe? A review of free web-accessible personal health record privacy policies. J Med Internet Res 2012;14(4):e114 [FREE Full text] [doi: 10.2196/jmir.1904] [Medline: 22917868]
16. Bajpai V, Gorthi R. On non-functional requirements: a survey. 2012 Presented at: Conf Electr Electron Comput Sci; 2012; Bhopal p. 1-4. [doi: 10.1109/SCEECS.2012.6184810]
17. Wazlawick RS. High-Level Requirements. In: Object-Oriented Analysis and Design for Information Systems. Amsterdam: Elsevier; 2014:29-57.
18. Welcome to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website!. 2015. URL: http://www.prisma-statement.org/ [accessed 2020-05-16]
19. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
20. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008;8:45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
21. World economic situation and prospects. New York: United Nations; 2020. URL: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2020_Annex.pdf [accessed 2020-06-15]

XSL•FO
RenderX

22. Tulu B, Trapp A, Strong D, Johnson S, Hoque M, Trudel J, et al. An analysis of patient portal utilization: what can we learn about online patient behavior by examining portal click data? Heal Syst 2017 Dec 19;5(1):66-79. [doi: 10.1057/hs.2015.5]
23. Iacona SL, Militello C, Serbanati L, Mastratisi M, Ricci F, Gilardi M. Personal health system: a tool to support the patient empowerment. 2015 Presented at: E-Health Bioeng Conf IEEE; 2015; Iasi p. 1-6. [doi: 10.1109/EHB.2015.7391370]
24. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with OpenNotes. Appl Clin Inform 2019 Jan;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
25. Abdulnabi M, Al-Haiqi A, Kiah MLM, Zaidan AA, Zaidan BB, Hussain M. A distributed framework for health information exchange using smartphone technologies. J Biomed Inform 2017 May;69:230-250 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.013] [Medline: 28433825]
26. Katehakis D, Kondylakis H, Koumakis L, Kouroubali A, Marias K. Integrated care solutions for the citizen: personal health record functional models to support interoperability. Eur J Biomed Informatics 2017;13(1):56.
27. Elston Lafata J, Miller CA, Shires DA, Dyer K, Ratliff SM, Schreiber M. Patients' adoption of and feature access within electronic patient portals. Am J Manag Care 2018 Nov 01;24(11):e352-e357 [FREE Full text] [Medline: 30452203]
28. Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P. Designing a patient portal for patient-centered care: cross-sectional survey. J Med Internet Res 2018 Oct 01;20(10):e269 [FREE Full text] [doi: 10.2196/jmir.9497] [Medline: 30287416]
29. Voigt I, Benedict M, Susky M, Scheplitz T, Frankowitz S, Kern R, et al. A digital patient portal for patients with multiple sclerosis. Front Neurol 2020;11:400 [FREE Full text] [doi: 10.3389/fneur.2020.00400] [Medline: 32670174]
30. Hill JN, Smith BM, Weaver FM, Nazi KM, Thomas FP, Goldstein B, et al. Potential of personal health record portals in the care of individuals with spinal cord injuries and disorders: provider perspectives. J Spinal Cord Med 2018 May;41(3):298-308 [FREE Full text] [doi: 10.1080/10790268.2017.1293760] [Medline: 28325112]
31. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
32. Rezaee R, AlizadehNaini M, Halim Z. Designing and implementation of web-based personal health record for patients with inflammatory bowel disease. Govaresh 2018;23(2):108.
33. Zhou L, DeAlmeida D, Parmanto B. Applying a user-centered approach to building a mobile personal health record app: development and usability study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]
34. Barbarito F, Pinciroli F, Barone A, Pizzo F, Ranza R, Mason J, et al. Implementing the lifelong personal health record in a regionalised health information system: the case of Lombardy, Italy. Comput Biol Med 2015 Apr;59:164-174. [doi: 10.1016/j.compbiomed.2013.10.021] [Medline: 24238969]
35. Clarke M, Schuetzler R, Windle J, Pachunka E, Fruhling A. Usability and cognitive load in the design of a personal health record. Heal Policy Technol 2020 Jun;9(2):218-224. [doi: 10.1016/j.hlpt.2019.10.002]
36. Javier SJ, Troszak LK, Shimada SL, McInnes DK, Ohl ME, Avoundjian T, et al. Racial and ethnic disparities in use of a personal health record by veterans living with HIV. J Am Med Inform Assoc 2019 Aug 01;26(8-9):696-702 [FREE Full text] [doi: 10.1093/jamia/ocz024] [Medline: 30924875]
37. Baudendistel I, Winkler E, Kamradt M, Längst G, Eckrich F, Heinze O, et al. Personal electronic health records: understanding user requirements and needs in chronic cancer care. J Med Internet Res 2015;17(5):e121 [FREE Full text] [doi: 10.2196/jmir.3884] [Medline: 25998006]
38. Song Y, Pak J, Kalabins A, Fouché S. Standard-based patient-centered personal health record system. Proc 11th Int Conf Ubiquitous Inf Manag Commun 2017:1-7. [doi: 10.1145/3022227.3022289]
39. Cernian A, Tiganoaia B, Sacala I, Pavel A, Iftemi A. PatientDataChain: a blockchain-based approach to integrate personal health records. Sensors (Basel) 2020 Nov 16;20(22):6538 [FREE Full text] [doi: 10.3390/s20226538] [Medline: 33207620]
40. Wright M, Flanagan M, Kunjan K, Doebbeling B, Toscos T. Missing links: Challenges in engaging the underserved with health information and communication technology. Proc 10th EAI Int Conf Pervasive Comput Technol Healthc 2016:122-129 [FREE Full text]
41. Minoletti S, Rapisarda R, Giraldo L, Grande M, Sommer J, Plazzotta F, et al. User-centered design of a pediatric vaccination module for patients. Stud Health Technol Inform 2019 Aug 21;264:1096-1100. [doi: 10.3233/SHTI190395] [Medline: 31438094]
42. Nazi KM, Turvey CL, Klein DM, Hogan TP. A decade of veteran voices: examining patient portal enhancements through the lens of user-centered design. J Med Internet Res 2018 Jul 10;20(7):e10413 [FREE Full text] [doi: 10.2196/10413] [Medline: 29991468]
43. Rief JJ, Hamm ME, Zickmund SL, Nikolajski C, Lesky D, Hess R, et al. Using health information technology to foster engagement: patients' experiences with an active patient health record. Health Commun 2017 Mar;32(3):310-319. [doi: 10.1080/10410236.2016.1138378] [Medline: 27223684]
44. Wang H, Miao X, Yang P. Design and implementation of personal health record systems based on knowledge graph. 2018 Presented at: 9th Int Conf Inf Technol Med Educ. IEEE; 2018; Hangzhou p. 133-136. [doi: 10.1109/ITME.2018.00039]

XSL•FO
RenderX

45. Dexheimer JW, Greiner MV, Beal SJ, Johnson D, Kachelmeyer A, Vaughn LM. Sharing personal health record data elements in protective custody: youth and stakeholder perspectives. J Am Med Inform Assoc 2019 Aug 01;26(8-9):714-721 [FREE Full text] [doi: 10.1093/jamia/ocz067] [Medline: 31365097]
46. Yen P, Walker DM, Smith JMG, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. Int J Med Inform 2018 Dec;110:10-18. [doi: 10.1016/j.ijmedinf.2017.11.007] [Medline: 29331248]
47. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
48. Smith MI, Garcia G, Simon M, Bruchanski L, Frangella J, Sommer JA, et al. Lessons learned after redesigning a personal health record. Stud Health Technol Inform 2017;245:216-220. [Medline: 29295085]
49. Avdagovska M, Ballermann M, Olson K, Graham T, Menon D, Stafinski T. Patient portal implementation and uptake: qualitative comparative case study. J Med Internet Res 2020 Jul 27;22(7):e18973 [FREE Full text] [doi: 10.2196/18973] [Medline: 32716308]
50. Carryer J, Kooienga S. Patients' experience and understanding of E-portals in rural general practice: an ethnographic exploration. J Prim Health Care 2017 Dec;9(4):262-268. [doi: 10.1071/HC17016] [Medline: 29530137]
51. Gee PM, Paterniti DA, Ward D, Soederberg MLM. e-Patients perceptions of using personal health records for self-management support of chronic illness. Comput Inform Nurs 2015 Jun;33(6):229-237. [doi: 10.1097/CIN.0000000000000151] [Medline: 25899440]
52. Robinson JR, Davis SE, Cronin RM, Jackson GP. Use of a patient portal during hospital admissions to surgical services. AMIA Annu Symp Proc 2016;2016:1967-1976 [FREE Full text] [Medline: 28269956]
53. Redelmeier DA, Kraus NC. Patterns in patient access and utilization of online medical records: analysis of MyChart. J Med Internet Res 2018 Feb 06;20(2):e43 [FREE Full text] [doi: 10.2196/jmir.8372] [Medline: 29410386]
54. Metting E, Schrage AJ, Kocks JW, Sanderman R, van der Molen T. Assessing the needs and perspectives of patients with asthma and chronic obstructive pulmonary disease on patient web portals: focus group study. JMIR Form Res 2018 Nov 22;2(2):e22 [FREE Full text] [doi: 10.2196/formative.8822] [Medline: 30684436]
55. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
56. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
57. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design considerations for patient portal adoption by low-income, older adults. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:3859-3868 [FREE Full text] [doi: 10.1145/2702123.2702392] [Medline: 27077140]
58. Jung SY, Lee K, Hwang H, Yoo S, Baek HY, Kim J. Support for sustainable use of personal health records: understanding the needs of users as a first step towards patient-driven mobile health. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e19 [FREE Full text] [doi: 10.2196/mhealth.6021] [Medline: 28232300]
59. Walker DM, Menser T, Yen P, McAlearney AS. Optimizing the user experience: identifying opportunities to improve use of an inpatient portal. Appl Clin Inform 2018 Dec;9(1):105-113. [doi: 10.1055/s-0037-1621732] [Medline: 29444536]
60. Geerts PAF, van der Weijden T, Loeffen PGM, Janssen LEF, Almekinders C, Wienhold TA, et al. Developing a patient portal for haematology patients requires involvement of all stakeholders and a customised design, tailored to the individual needs. BMC Med Inform Decis Mak 2019 Jul 11;19(1):129 [FREE Full text] [doi: 10.1186/s12911-019-0868-y] [Medline: 31296210]
61. Strudwick G, Booth RG, McLean D, Leung K, Rossetti S, McCann M, et al. Identifying indicators of meaningful patient portal use by psychiatric populations. Inform Health Soc Care 2020 Oct 01;45(4):396-409. [doi: 10.1080/17538157.2020.1776291] [Medline: 32603617]
62. Tarver WL, Robb BW, Haggstrom DA. Usefulness and usability of a personal health record and survivorship care plan for colorectal cancer survivors: survey study. JMIR Cancer 2019 Aug 20;5(2):e10692 [FREE Full text] [doi: 10.2196/10692] [Medline: 31432780]
63. Subbe CP, Øvretveit J, Quinn N, Wyatt JC. Digital technology: opportunities and barriers for usage of personal health records in hospital. Future Healthc J 2019 Feb;6(1):52-56 [FREE Full text] [doi: 10.7861/futurehosp.6-1-52] [Medline: 31098587]
64. Wildenbos GA, Maasri K, Jaspers M, Peute L. Older adults using a patient portal: registration and experiences, one year after implementation. Digit Health 2018;4:2055207618797883 [FREE Full text] [doi: 10.1177/2055207618797883] [Medline: 30186619]
65. Hanna L, Gill SD, Newstead L, Hawkins M, Osborne RH. Patient perspectives on a personally controlled electronic health record used in regional Australia. Health Inf Manag 2017 Jan;46(1):42-48. [doi: 10.1177/1833358316661063] [Medline: 27486184]

XSL•FO
RenderX

66. Proctor JW, Lewis MG, Mansell PW. Immunotherapy for cancer: an overview. Can J Surg 1976 Jan;19(1):12-19. [doi: 10.1145/3290688.3290727]
67. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
68. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019 Jun;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
69. Wells S, Mahony F, Huang Y, Day K. Perspectives of New Zealand patients and GPs at the beginning of patient portal implementation. J Prim Health Care 2019;11(4):315. [doi: 10.1071/HC19016]
70. Gupta S, Kaur N, Jain D. mSwasthya: A mobile-enabled personal health record management system. 2015 Presented at: Int Conf Comput Commun Autom; 2015; Greater Noida p. 374-379. [doi: 10.1109/CCAA.2015.7148404]
71. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
72. Aliakbarpoor Y, Comai S, Pozzi G. Designing a HL7 compatible personal health record for mobile devices. 2017 Presented at: IEEE 3rd Int Forum Res Technol Soc Ind; 2017; Modena p. 1-6. [doi: 10.1109/RTSI.2017.8065881]
73. Brandner A, Schreiweis B, Aguduri LS, Bronsch T, Kunz A, Pensold P, et al. The patient portal of the personal cross-enterprise electronic health record (PEHR) in the Rhine-Neckar-Region. Stud Health Technol Inform 2016;228:157-161. [Medline: 27577362]
74. O'Leary KJ, Sharma RK, Killarney A, O'Hara LS, Lohman ME, Culver E, et al. Patients' and healthcare providers' perceptions of a mobile portal application for hospitalized patients. BMC Med Inform Decis Mak 2016 Sep 21;16(1):123 [FREE Full text] [doi: 10.1186/s12911-016-0363-7] [Medline: 27653854]
75. Brown SM, Bell SK, Roche SD, Dente E, Mueller A, Kim T, et al. Preferences of current and potential patients and family members regarding implementation of electronic communication portals in intensive care units. Ann Am Thorac Soc 2016 Mar;13(3):391-400. [doi: 10.1513/AnnalsATS.201509-638OC] [Medline: 26700656]
76. Marchak JG, Cherven B, Williamson Lewis R, Edwards P, Meacham LR, Palgon M, et al. User-centered design and enhancement of an electronic personal health record to support survivors of pediatric cancers. Support Care Cancer 2020 Aug;28(8):3905-3914 [FREE Full text] [doi: 10.1007/s00520-019-05199-w] [Medline: 31853699]
77. Woollen J, Prey J, Wilcox L, Sackeim A, Restaino S, Raza ST, et al. Patient experiences using an inpatient personal health record. Appl Clin Inform 2016;7(2):446-460 [FREE Full text] [doi: 10.4338/ACI-2015-10-RA-0130] [Medline: 27437053]
78. Eschler J, Meas PL, Lozano P, McClure JB, Ralston JD, Pratt W. Integrating the patient portal into the health management work ecosystem: user acceptance of a novel prototype. AMIA Annu Symp Proc 2016;2016:541-550 [FREE Full text] [Medline: 28269850]
79. Fosser SM, Sommer J, Simón M, Giraldo L, Plazzotta F, Luna D. User-centered design of a patient medication reconciliation module in an integrated personal health record. Stud Health Technol Inform 2019 Aug 21;264:1278-1282. [doi: 10.3233/SHTI190432] [Medline: 31438131]
80. Wilcox L, Woollen J, Prey J, Restaino S, Bakken S, Feiner S, et al. Interactive tools for inpatient medication tracking: a multi-phase study with cardiothoracic surgery patients. J Am Med Inform Assoc 2016 Jan;23(1):144-158 [FREE Full text] [doi: 10.1093/jamia/ocv160] [Medline: 26744489]
81. Conway NT, Allardice B, Wake DJ, Cunningham SG. User experiences of an electronic personal health record for diabetes. J Diabetes Sci Technol 2019 Jul;13(4):744-750 [FREE Full text] [doi: 10.1177/1932296818818837] [Medline: 30585085]
82. Schladen MM, Ljungberg IH, Underwood JA, Groah SL. Exploration of the personal health record as a tool for spinal cord injury health self-management and coordination of care. Top Spinal Cord Inj Rehabil 2017;23(3):218-225 [FREE Full text] [doi: 10.1310/sci2303-218] [Medline: 29339897]
83. Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, et al. Developing a shared patient-centered, web-based medication platform for type 2 diabetes patients and their health care providers: qualitative study on user requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]
84. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
85. Park YR, Lee Y, Kim JY, Kim J, Kim HR, Kim Y, et al. Managing patient-generated health data through mobile personal health records: analysis of usage data. JMIR Mhealth Uhealth 2018 Apr 09;6(4):e89 [FREE Full text] [doi: 10.2196/mhealth.9620] [Medline: 29631989]
86. Cabitza F, Simone C, De Michelis G. User-driven prioritization of features for a prospective InterPersonal Health Record: perceptions from the Italian context. Comput Biol Med 2015 Apr;59:202-210. [doi: 10.1016/j.compbiomed.2014.03.009] [Medline: 24768267]

XSL•FO
RenderX

87. Davis S, MacKay L. Moving beyond the rhetoric of shared decision-making: designing personal health record technology with young adults with type 1 diabetes. Can J Diabetes 2020 Jul;44(5):434-441. [doi: 10.1016/j.jcjd.2020.03.009] [Medline: 32616277]
88. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. BMC Health Serv Res 2018 Jun 14;18(1):445 [FREE Full text] [doi: 10.1186/s12913-018-3254-z] [Medline: 29898730]
89. Lalitaphanit K, Theeraroungchaisri A. Factors affecting community pharmacy customers? Decision to use personal health records via smartphone. Thai J Pharm Sci (Supplement Issue) 2016;40:167 [FREE Full text]
90. Sieverink F, Kelders S, Braakman-Jansen A, van Gemert-Pijnen J. Evaluating the implementation of a personal health record for chronic primary and secondary care: a mixed methods approach. BMC Med Inform Decis Mak 2019 Nov 27;19(1):241 [FREE Full text] [doi: 10.1186/s12911-019-0969-7] [Medline: 31775734]
91. Samet S, Ishraque M, Sharma A. Privacy-preserving personal health record (P3HR). Proc 7th Int Conf Softw Inf Eng 2018:22-26. [doi: 10.1145/3220267.3220271]
92. Turner AM, Osterhage K, Hartzler A, Joe J, Lin L, Kanagat N, et al. Use of patient portals for personal health information management: the older adult perspective. AMIA Annu Symp Proc 2015;2015:1234-1241 [FREE Full text] [Medline: 26958263]
93. Khaneghah PA, Miguel-Cruz A, Bentley P, Liu L, Stroulia E, Ferguson-Pell M. Users' attitudes towards personal health records: a cross-sectional pilot study. Appl Clin Inform 2016;7(2):573-586 [FREE Full text] [doi: 10.4338/ACI-2015-12-RA-0180] [Medline: 27437062]
94. Forchuk C, Reiss JP, O'Regan T, Ethridge P, Donelle L, Rudnick A. Client perceptions of the mental health engagement network: a qualitative analysis of an electronic personal health record. BMC Psychiatry 2015 Oct 14;15:250 [FREE Full text] [doi: 10.1186/s12888-015-0614-7] [Medline: 26467210]
95. Marthick M, Janssen A, Cheema BS, Alison J, Shaw T, Dhillon H. Feasibility of an interactive patient portal for monitoring physical activity, remote symptom reporting, and patient education in oncology: qualitative study. JMIR Cancer 2019 Nov 28;5(2):e15539 [FREE Full text] [doi: 10.2196/15539] [Medline: 31778123]
96. Sheehan B, Lucero RJ. Initial usability and feasibility evaluation of a personal health record-based self-management system for older adults. EGEMS (Wash DC) 2015;3(2):1152 [FREE Full text] [doi: 10.13063/2327-9214.1152] [Medline: 26290889]
97. Arcia A. Time to push: use of gestational age in the electronic health record to support delivery of relevant prenatal education content. EGEMS (Wash DC) 2017 Apr 20;5(2):5 [FREE Full text] [doi: 10.13063/2327-9214.1281] [Medline: 29881749]
98. Fonseca F, Peixoto H, Braga J, Machado J, Abelha A. Smart mobile computing in pregnancy care. Proc 34th Int Conf Comput Their Appl 2019:209-212. [doi: 10.29007/sr6q]
99. Iatraki G, Kondylakis H, Koumakis L, Chatzimina M, Kazantzaki E, Marias K, et al. Personal Health Information Recommender: implementing a tool for the empowerment of cancer patients. Ecancermedicalscience 2018;12:851 [FREE Full text] [doi: 10.3332/ecancer.2018.851] [Medline: 30079113]
100. van den Heuvel SCGH, Meije D, Regeer EJ, Sinnema H, Riemersma RF, Kupka RW. The user experiences and clinical outcomes of an online personal health record to support self-management of bipolar disorder: A pretest-posttest pilot study. J Affect Disord 2018 Oct 01;238:261-268. [doi: 10.1016/j.jad.2018.05.069] [Medline: 29894931]
101. Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of myhealthkeeper. J Med Internet Res 2017 Dec 07;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]
102. Cronin RM, Conway D, Condon D, Jerome RN, Byrne DW, Harris PA. Patient and healthcare provider views on a patient-reported outcomes portal. J Am Med Inform Assoc 2018 Nov 01;25(11):1470-1480. [doi: 10.1093/jamia/ocy111] [Medline: 30239733]
103. Fuji KT, Abbott AA, Galt KA. A qualitative study of how patients with type 2 diabetes use an electronic stand-alone personal health record. Telemed J E Health 2015 Apr;21(4):296-300. [doi: 10.1089/tmj.2014.0084] [Medline: 25614996]
104. Ploner N, Neurath MF, Schoenthaler M, Zielke A, Prokosch H. Concept to gain trust for a German personal health record system using public cloud and FHIR. J Biomed Inform 2019 Jul;95:103212 [FREE Full text] [doi: 10.1016/j.jbi.2019.103212] [Medline: 31112761]
105. Farinango CD, Benavides JS, Cerón JD, López DM, Álvarez RE. Human-centered design of a personal health record system for metabolic syndrome management based on the ISO 9241-210:2010 standard. J Multidiscip Healthc 2018;11:21-37 [FREE Full text] [doi: 10.2147/JMDH.S150976] [Medline: 29386903]
106. Pohlmann S, Kunz A, Ose D, Winkler EC, Brandner A, Poss-Doering R, et al. Digitalizing health services by implementing a personal electronic health record in Germany: qualitative analysis of fundamental prerequisites from the perspective of selected experts. J Med Internet Res 2020 Jan 29;22(1):e15102 [FREE Full text] [doi: 10.2196/15102] [Medline: 32012060]
107. Roehrs A, da Costa CA, da Rosa RR. OmniPHR: a distributed architecture model to integrate personal health records. J Biomed Inform 2017 Dec;71:70-81 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.012] [Medline: 28545835]
108. Jing F. An empirical study on the features influencing users? Adoption towards personal health records system. 13th Int Conf Serv Syst Serv Manag 2016:1-6. [doi: 10.1109/icsssm.2016.7538554]

XSL•FO
RenderX

109. Poss-Doering R, Kunz A, Pohlmann S, Hofmann H, Kiel M, Winkler EC, et al. Utilizing a prototype patient-controlled electronic health record in Germany: qualitative analysis of user-reported perceptions and perspectives. JMIR Form Res 2018 Aug 03;2(2):e10411 [FREE Full text] [doi: 10.2196/10411] [Medline: 30684411]
110. Gagnon M, Payne-Gagnon J, Breton E, Fortin J, Khoury L, Dolovich L, et al. Adoption of electronic personal health records in Canada: perceptions of stakeholders. Int J Health Policy Manag 2016 Jul 01;5(7):425-433 [FREE Full text] [doi: 10.15171/ijhpm.2016.36] [Medline: 27694670]
111. Ruhi U, Majedi A. Int Conf e-Health 2015 2015:43-54 [FREE Full text]
112. Urbauer P, Sauermann S, Frohner M, Forjan M, Pohn B, Mense A. Applicability of IHE/Continua components for PHR systems: learning from experiences. Comput Biol Med 2015 Apr;59:186-193. [doi: 10.1016/j.compbiomed.2013.12.003] [Medline: 24374230]
113. Baudendistel I, Winkler EC, Kamradt M, Brophy S, Längst G, Eckrich F, et al. Cross-sectoral cancer care: views from patients and health care professionals regarding a personal electronic health record. Eur J Cancer Care (Engl) 2016 Feb 03;26(2):e12429. [doi: 10.1111/ecc.12429] [Medline: 26840784]
114. Vimalachandran P, Liu H, Lin Y, Ji K, Wang H, Zhang Y. Improving accessibility of the Australian My Health Records while preserving privacy and security of the system. Health Inf Sci Syst 2020 Dec;8(1):31. [doi: 10.1007/s13755-020-00126-4] [Medline: 33088487]
115. Wang T, Dolezel D. Usability of web-based personal health records: an analysis of consumers' perspectives. Perspect Health Inf Manag 2016;13:1f [FREE Full text] [Medline: 27134611]
116. Pang P, McKay D, Chang S, Chen Q, Zhang X, Cui L. Privacy concerns of the Australian My Health Record: implications for other large-scale opt-out personal health records. Inform Process Manag 2020 Nov;57(6):102364. [doi: 10.1016/j.ipm.2020.102364]
117. Collins SA, Rozenblum R, Leung WY, Morrison CR, Stade DL, McNally K, et al. Acute care patient portals: a qualitative study of stakeholder perspectives on current practices. J Am Med Inform Assoc 2017 Apr 01;24(e1):e9-e17. [doi: 10.1093/jamia/ocw081] [Medline: 27357830]
118. Fernando M, Fidge C, Sahama T. Design guidelines for effective occupation-based personal health records. Proc Australas Comput Sci Week Multiconf 2020:1-10. [doi: 10.1145/3373017.3373042]
119. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in england: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
120. Alpert JM, Krist AH, Aycock RA, Kreps GL. Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. J Med Internet Res 2016 May 17;18(5):e112 [FREE Full text] [doi: 10.2196/jmir.5451] [Medline: 27188953]
121. Rau H, Wu Y, Chu C, Wang F, Hsu M, Chang C, et al. Importance-performance analysis of personal health records in Taiwan: a web-based survey. J Med Internet Res 2017 Apr 27;19(4):e131 [FREE Full text] [doi: 10.2196/jmir.7065] [Medline: 28450273]
122. Abdekhoda M, Dehnad A, Khezri H. The effect of confidentiality and privacy concerns on adoption of personal health record from patient’s perspective. Health Technol 2019 Jan 6;9(4):463-469. [doi: 10.1007/s12553-018-00287-z]
123. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: a qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 04;23(2):848 [FREE Full text] [Medline: 27869582]
124. Tao D, Yuan J, Shao F, Li D, Zhou Q, Qu X. Factors affecting consumer acceptance of an online health information portal among young internet users. Comput Inform Nurs 2018 Nov;36(11):530-539. [doi: 10.1097/CIN.0000000000000467] [Medline: 30074918]
125. Vassilakopoulou P, Grisot M, Aanestad M. Between personal and common: the design of hybrid information spaces. Comput Supported Coop Work 2018 Jan 8;27(3-6):1085-1112. [doi: 10.1007/s10606-017-9304-y]
126. Walsh L, Hemsley B, Allan M, Adams N, Balandin S, Georgiou A, et al. The e-Health literacy demands of Australia's My Health Record: a heuristic evaluation of usability. Perspect Health Inf Manag 2017;14(Fall):1f [FREE Full text] [Medline: 29118683]
127. Barnsteiner J. Medication reconciliation. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville: Agency for Healthcare Research and Quality; 2008:459-472.
128. Triggle N. Covid-19: Surge leaves key hospital services "in crisis". BBC. 2021 Jan 14. URL: https://www.bbc.com/news/health-55575112 [accessed 2021-02-09]
129. Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, et al. Health observation app for COVID-19 symptom tracking integrated with personal health records: proof of concept and practical use study. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e19902. [doi: 10.2196/19902] [Medline: 32568728]
130. HL7 Personal Health Record System Functional Model, Release 1. Health Level Seven International. 2014. URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=88 [accessed 2021-04-20]
131. Personal Health Record System Functional Model (PHR-S FM), R1. Health Level Seven International. 2014. URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=88 [accessed 2021-04-20]

XSL•FO
RenderX

132. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]

Abbreviations

EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level 7
HL7 PHR-S FM: Health Level 7 Personal Health Record System Functional Model
IN: information infrastructure
PH: personal health
PHR: personal health record
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-analyses
S: supportive

Edited by R Kukafka; submitted 09.12.20; peer-reviewed by S Davis, H Mehdizadeh, T Ndabu, PCI Pang; comments to author 16.01.21; revised version received 07.03.21; accepted 24.05.21; published 21.07.21.

Please cite as:
Harahap NC, Handayani PW, Hidayanto AN
Functionalities and Issues in the Implementation of Personal Health Records: Systematic Review
J Med Internet Res 2021;23(7):e26236
URL: https://www.jmir.org/2021/7/e26236
doi:10.2196/26236
PMID:34287210

©Nabila Clydea Harahap, Putu Wuri Handayani, Achmad Nizar Hidayanto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)–Based Data Model: Development and Topic Modeling Study

Amrita De[1*], MA; Ming Huang[1*], PhD; Tinghao Feng[2], MSc; Xiaomeng Yue[3], PhD; Lixia Yao[1], PhD

[1]Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, United States

[2]Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, United States

[3]Division of Pharmacy Practice and Administrative Sciences, James L. Winkle College of Pharmacy, University of Cincinnati, Cincinnati, OH, United States

[*]these authors contributed equally

Corresponding Author:
Lixia Yao, PhD
Department of Artificial Intelligence and Informatics
Mayo Clinic
200 1st St SW
Rochester, MN
United States
Phone: 1 5072842511
Email: lixia.cn.yao@gmail.com

Abstract

Background: Patient portals tethered to electronic health records systems have become attractive web platforms since the enacting of the Medicare Access and Children's Health Insurance Program Reauthorization Act and the introduction of the *Meaningful Use* program in the United States. Patients can conveniently access their health records and seek consultation from providers through secure web portals. With increasing adoption and patient engagement, the volume of patient secure messages has risen substantially, which opens up new research and development opportunities for patient-centered care.

Objective: This study aims to develop a data model for patient secure messages based on the Fast Healthcare Interoperability Resources (FHIR) standard to identify and extract significant information.

Methods: We initiated the first draft of the data model by analyzing FHIR and manually reviewing 100 sentences randomly sampled from more than 2 million patient-generated secure messages obtained from the online patient portal at the Mayo Clinic Rochester between February 18, 2010, and December 31, 2017. We then annotated additional sets of 100 randomly selected sentences using the Multi-purpose Annotation Environment tool and updated the data model and annotation guideline iteratively until the interannotator agreement was satisfactory. We then created a larger corpus by annotating 1200 randomly selected sentences and calculated the frequency of the identified medical concepts in these sentences. Finally, we performed topic modeling analysis to learn the hidden topics of patient secure messages related to 3 highly mentioned microconcepts, namely, fatigue, prednisone, and patient visit, and to evaluate the proposed data model independently.

Results: The proposed data model has a 3-level hierarchical structure of health system concepts, including 3 macroconcepts, 28 mesoconcepts, and 85 microconcepts. Foundation and base macroconcepts comprise 33.99% (841/2474), clinical macroconcepts comprise 64.38% (1593/2474), and financial macroconcepts comprise 1.61% (40/2474) of the annotated corpus. The top 3 mesoconcepts among the 28 mesoconcepts are condition (505/2474, 20.41%), medication (424/2474, 17.13%), and practitioner (243/2474, 9.82%). Topic modeling identified hidden topics of patient secure messages related to fatigue, prednisone, and patient visit. A total of 89.2% (107/120) of the top-ranked topic keywords are actually the health concepts of the data model.

Conclusions: Our data model and annotated corpus enable us to identify and understand important medical concepts in patient secure messages and prepare us for further natural language processing analysis of such free texts. The data model could be potentially used to automatically identify other types of patient narratives, such as those in various social media and patient forums. In the future, we plan to develop a machine learning and natural language processing solution to enable automatic triaging solutions to reduce the workload of clinicians and perform more granular content analysis to understand patients' needs and improve patient-centered care.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e26770) doi:10.2196/26770

KEYWORDS

patient secure messages; patient portal; data model; FHIR; annotated corpus; topic modeling

Introduction

Background

In the United States, the Medicare Access and CHIP (Children's Health Insurance Program) Reauthorization Act [1] and *Meaningful Use* program [2] have incentivized the growing adoption of electronic health records (EHRs) and patient health records with the goal of improving the quality of health care delivery systems. Consequently, many health care delivery systems now offer patient portals, tethered to their EHR systems, that allow patients to access their medical records and communicate with their clinicians through secure messages [3]. Patient portals encourage patients to become equal partners in their care and health management and be more engaged and participatory in shared decision making [4]. After the Health Information Technology for Economic and Clinical Health Act was enacted in 2009, patient portals have gained widespread adoption by health care delivery systems in the United States [5,6]. In 2017, more than 90% of the health care delivery systems, including the Veterans Health Administration, Mass General Brigham, Kaiser Permanente, and the Mayo Clinic, offered patient portal access to their patients [7]. Currently, patients send secure web-based messages to request medical appointments and prescription refills [8,9]. Clinicians send patients appointment reminders and promote timely preventive care [10,11]. Patients and clinicians can communicate back and forth easily and in a timely manner about complex situations such as new symptoms, follow-up visits, medication concerns, and medical questions.

With the increase in the number of patients signing up for these portals, the number of secure messages has risen substantially [12-15]. Unfortunately, the content of the large number of patient secure messages in free-text format has not been processed and analyzed systematically and incorporated into the present EHR systems organically to unfold its potential for improving patient-centered care because of technical hurdles. For instance, existing annotated corpora have been mainly developed for sublanguages such as scientific literature in biomedicine and clinical notes; no annotated corpus is available for developing natural language processing (NLP) capabilities in patient-generated formal language. In this study, we propose to develop a data model of health concepts for patient secure messages based on the Fast Healthcare Interoperability Resources (FHIR) standard [16].

A data model is usually made up of entities that represent important items in the domain and relationship assertions among the entities. In our case, a data model will illustrate the key concepts occurring in patient secure messages and the relationships among them. The data model is critical to the development of any information system (eg, a health information exchange system or NLP-based semantic representation system for a patient portal) by providing the definition of the concepts and format of data. Building the factual and useful data model requires a deep understanding of the underlying process and data. Therefore, we also create a large annotated corpus for analyzing the contents of sampled patient secure messages to better understand patients' concerns. Once complete, we further apply topic modeling techniques independently to investigate whether the patients' focuses and concerns in 3 common medical conditions align with the developed data model. We build the annotated corpus primarily to build the data model, and topic modeling can serve as an independent and primitive validation of the data model. We choose topic modeling instead of information extraction because we build the annotated corpus primarily to build the data model. Topic modeling, as an unsupervised method, generates results independently of the corpus and thus can serve as an independent validation of the data model. We expect a much more rigorous evaluation and validation by building collaborations and partnerships with domestic and international researchers in the field.

With all the necessary preparation, our ultimate objective is to develop an NLP system that will automatically identify and extract significant information from unstructured patient secure messages for the purpose of automatically triaging patient secure messages, reducing the workload of clinicians by chatbot, and performing more granular and sophisticated content analysis to understand patients' needs and improve shared decision making and patient-centered care.

Related Work

As patient secure messages are relatively new, very little research has focused on automatically identifying and standardizing their content despite their important implications. North et al [17] analyzed the content of 6430 secure messages to assess the overall risk associated with the messages and to determine whether patients were using portal messages for symptoms requiring urgent evaluation. Their study showed that patients used portal messages 3.5% of the time for potentially high-risk symptoms of chest pain, breathing concerns, abdominal pain, palpitations, lightheadedness, and vomiting. Sulieman et al [18] also developed machine learning models on patient portal secure messages regarding surgical issues to identify message threads that involve medical decision making from a health care provider and to classify the complexity of the decision. Cronin et al [19] built patient portal message classifiers using rule-based and NLP techniques such as the bag-of-words model. They curated a gold standard data set of 3253 portal messages annotated by communication types such as informational, medical, logistical, and social. This study also focuses on developing a data model—a standard framework—to address the issues of content analysis, information extraction, and integration of significant information from patient secure messages leveraging Health Level-7 (HL7) FHIR.

HL7 is a nonprofit standard development organization accredited by the American National Standards Institute, and it is dedicated to providing a comprehensive framework and related standards

XSL•FO
RenderX

for the data exchange, integration, sharing, and retrieval of electronic health information [20]. FHIR is an improved health data exchange standard that comprehensively defines how information can be exchanged among different systems regardless of how it has been stored and allows health care information to be accessible to those who need it for the benefit of health care quickly and easily [21]. Instead of traditional document-centric approaches, HL7 FHIR takes a modular approach and represents atomic or granular health care data (eg, heart rate, procedure, medication, and allergies) as independent modular entities, concepts, and actions involved in health care information analysis, exchange, and integration as resources [22]. Existing studies on developing data standards have mostly focused on analyzing, extracting, and integrating structured data from EHRs, mobile-based patient health records, and medical apps. In 2018, Hong et al [23] for the first time introduced a scalable and standard-based framework for analyzing and integrating both structured and unstructured EHR data by leveraging the FHIR specification [23]. The scope and use of the FHIR framework do not completely meet the requirements of our study. We aim to develop an HL7 FHIR–based data model that precisely analyzes and extracts patient secure messages.

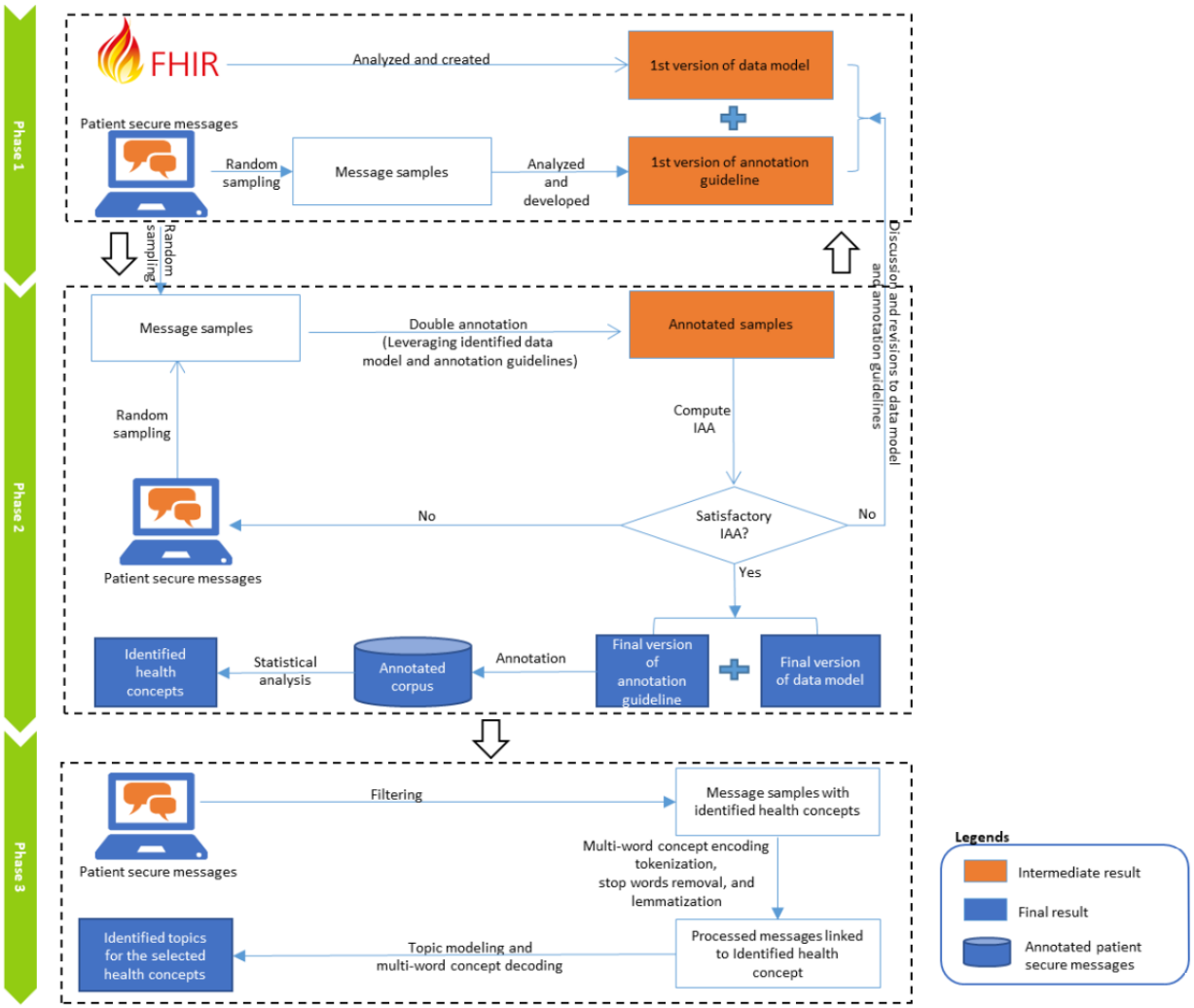
Methods

Overview

A 3-phase workflow for developing the data model and annotated corpus is shown in Figure 1. We collected more than 2 million patient-generated secure messages from Mayo Clinic Patient Online Services [24]. We developed the first draft of the data model and annotation guideline by analyzing FHIR and manually reviewing 100 sentences from the sampled secure messages. We then randomly selected, annotated, and examined additional sets of 100 sentences from the secure messages to iteratively update the data model and annotation guideline until interannotator agreement (IAA) was achieved. Subsequently, we created the annotated corpus by annotating 1200 sentences from the randomly selected 2100 sentences. Finally, we calculated the frequency of the identified health concepts in the annotated corpus and performed topic modeling to extract hidden topics of all the patient secure messages linked to frequently mentioned health concepts. In the following sections, we will discuss data collection and preprocessing, design of the data model, development of the annotation guideline, creation and analysis of the annotated corpus, and topic modeling in more detail. The entire data model can be found in Multimedia Appendix 1. The details of data set collection and processing and annotation and topic modeling and annotation guideline can be found in Multimedia Appendix 2 [16,24-33] and Multimedia Appendix 3 [1-3,16,24].

XSL•FO
RenderX

Figure 1. The workflow for developing the data model for patient secure messages, annotated corpus, and topic modeling analysis. FHIR: Fast Healthcare Interoperability Resources; IAA: interannotator agreement.



Ethics Approval

No patient was exposed to any intervention. We used data from the Mayo Clinic Unified Data Platform to develop the annotated corpus and for the analysis. The study was approved by the Mayo Clinic Institutional Review Board (19-002211).

Data Set Collection and Preprocessing

The Mayo Clinic Rochester started the patient portal (Patient Online Services) in 2010 for primary care practice and later extended it to specialty practice in 2013 [24]. We collected more than 2 million patient-generated secure messages from the online patient portal between February 18, 2010, and December 31, 2017. We removed messages with empty message bodies. Each message has a unique message ID, previous message ID, initial message ID, sender ID, recipient ID, date and time of the message creation, message subject, and message body. As we have mentioned earlier, the details are in Multimedia Appendix 2.

Design of a Data Model for Patient Secure Messages

Creating a data model for unstructured patient narratives is a very challenging task. After the literature review and initial analysis of the sampled secure messages, we decided to use the FHIR standard to develop the data model (Multimedia Appendix 1) because it comprehensively represents the modular entities, concepts, and actions involved in health care information exchange.

FHIR defines a hierarchical set of core and infrastructure resources for handling health concepts in an EHR [34]. After analyzing version 4 of the FHIR standard [16] and the sampled secure messages, we generated the first draft of the data model with 3 hierarchical levels—macroconcepts, mesoconcepts, and microconcepts—to extract information from the patient secure messages. We merged and revised some concepts from FHIR after analyzing the messages. For example, we merged 2 similar but separately defined top-level concepts (foundation concept and base concept) in FHIR into 1 macroconcept (foundation and base concept) in the data model. We also introduced a microconcept—*unspecified*—as an attribute under all mesoconcepts. *Unspecified* refers to the general terms under most of the mesoconcepts that cannot be categorized into any specific microconcepts. We deleted some mesoconcepts under clinical macroconcepts such as *clinical impression*, *detected issue*, *medication knowledge*, *molecular sequence*, and *care team* because they are not relevant to patient secure messages. The data model also underwent rounds of revisions during the annotation process to handle inconsistencies, and disagreement

XSL•FO
RenderX

occurred between the annotators. For instance, in the first draft of the data model, the mesoconcept *patient* has 7 microconcepts following FHIR, such as *identifier*, *name*, *telecom*, *gender*, *birthdate*, *address*, and *marital status*. In the final data model, the mesoconcept *patient* has 4 microconcepts, such as *unspecified*, *privacy*, *lifestyle*, and *diet*, to better understand patients' medical records and history. All personal information related to the patient is kept under *privacy* to maintain data privacy.

Development of an Annotation Guideline

For any annotation task, it is important that all annotators follow the same standard (annotation guideline) to minimize annotation confusion and errors. We manually reviewed several sets of 100 randomly selected message sentences to develop an annotation guideline. More specifically, we created the first version of our annotation guideline by analyzing the first set of 100 message sentences. Subsequently, our annotators independently annotated another set of 100 message sentences to examine the effectiveness of the annotation guideline until the annotators reached a considerable amount of agreement.

The final annotation guideline consists of 2 sections: (1) the first section discusses the general annotation rules; (2) the second section describes the health concepts and associative rules for the identification and extraction of these concepts together with specific examples. For instance, it is challenging to differentiate 2 microconcepts (*name* and *symptoms*) under the mesoconcept *condition*. The *condition-name* microconcept refers to the name of a disease and/or medical condition (eg, *rheumatoid arthritis*, *diabetes mellitus*, and *influenza*). The *condition-symptom* microconcept denotes a physical or mental feature or symptom (eg, *sore bottom* and *numb arm*) of a disease and/or medical condition. As per general annotation rules, all modifiers (eg, adjectives and adverbs) and possessives are removed for annotation to make text spans consistent. For the cases *special calcium pill*, *this medicine*, and *my rheumatoid arthritis*, the modifiers and possessives *special*, *this*, and *my* are not considered for annotation. All patient private information (eg, name, identity number, and clinic number) is not disclosed to maintain data privacy. The guideline clearly defines the concepts and rules to lessen the scope of ambiguity and error in the annotation and increase the possibility of agreement among the annotators to develop a quality corpus. The complete guideline for annotating patient secure messages is listed in Multimedia Appendix 3.

Development and Analysis of an Annotated Corpus of Patient Secure Messages

We chose the Multi-purpose Annotation Environment tool to annotate patient secure messages because of its ease of use and ability to fix discrepancies [25]. The Multi-purpose Annotation Environment tool requires a document type definition file with concept tags and attributes. We chose 2 professional annotators—a clinically trained linguist and a student pharmacist—to create a standard error-free corpus. They initially analyzed and revised several sets of 100 randomly sampled message sentences to iteratively improve the data model and annotation guideline. To check the consistency between the annotators, we calculated the F1 score of 2 separate annotation sets as the IAA score using General Architecture for Text Engineering software [26]. The IAA scores were computed using the F1 score as a criterion at the level of entities. This helped us to understand the span of the concepts on which the annotators agreed, disagreed, and partially agreed. We decided to follow the lenient parameter for measuring the IAA. With the lenient approach, the annotations that overlap are counted as a partial match, in contrast with the strict approach in which the annotations have to match with one another completely. We can consider this sentence as an example: "My mother has severe sinus headaches for several months." Now, if one annotator annotates "sinus headache" and another annotates "severe sinus headache," then the strict IAA approach will give us no match, but the lenient approach will consider this as an overlap. The F1 scores of the first set of annotations were 0.42 (macro mean) and 0.67 (micro mean). After discussing and resolving the disagreements, the annotators annotated another set of the same sentences, and the F1 scores were quite satisfactory: 0.62 (macro mean) and 0.74 (micro mean). Our annotators finally developed a quality corpus of 1200 randomly selected sentences.

After annotation, we performed summary statistics to calculate the frequencies of the identified health concepts (ie, macroconcepts, mesoconcepts, and microconcepts) in the annotated corpus of patient secure messages. The distribution of these health concepts helps us to understand which concepts patients were mostly concerned about and communicated to their health care providers. The details are provided in Multimedia Appendix 2.

Topic Modeling

After analyzing the annotated corpus, we selected 3 health microconcepts (ie, fatigue, prednisone, and patient visit) as representative cases for topic modeling analysis. The chosen microconcepts and the corresponding meso concepts and macroconcepts were frequently discussed in the patient portal messages (refer to the *Results* section for more details). Fatigue is an instance of a top-mentioned microconcept, symptom, under the condition mesoconcept and clinical macroconcept. Prednisone is a case of a microconcept, name, under the medication mesoconcept and clinical macroconcept, about which patients have expressed most concern in the patient secure messages. Patient visit is an example of a largely discussed microconcept, type, under the appointment mesoconcept and the foundation and base macroconcept.

After multi-word concept encoding and health concept recognition using MetaMap [27], we collected 41,490, 27,743, and 95,533 patient secure messages that mentioned the health microconcepts fatigue, prednisone, and patient visit, respectively, to examine the focus of those messages. MetaMap is a highly configurable program. It has been developed by Dr Alan Aronson at the National Library of Medicine to map biomedical texts to the Unified Medical Language System.

Topic modeling automatically identifies topics or themes in a large collection of documents in terms of a set of keywords that occur together and most frequently [35,36]. We used latent Dirichlet allocation (LDA) [28], a state-of-the-art unsupervised topic modeling method as implemented in Machine Learning

XSL•FO
RenderX

for Language Toolkit [29], to learn the hidden topics of patient secure messages related to each of the 3 health microconcepts. After tokenization, stop word removal, and lemmatization, each patient secure message was converted into a vocabulary vector where the elements were the frequency of each lemma (including Concept Unique Identifier encoded by MetaMap) without considering the order of the lemma.

We quantitatively calculated topic coherence [30] and asked domain experts to qualitatively evaluate the learned topics. More specifically, we evaluated the topic coherence at different topic numbers (ie, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30) to determine the optimal topic number for the 3 selected health microconcepts. We found that the optimal topic number was 8 for fatigue-related messages, 10 for prednisone-related messages, and 12 for patient visit–related messages. We also investigated the hyperparameters of and in the LDA. controls the topic distributions over a document. A smaller results in fewer topics that are statistically associated with a document. determines the word distributions over a topic. A smaller leads to fewer words that are statistically linked to a topic. We set the automatic optimization of the hyperparameters of and in Machine Learning for Language Toolkit (Mallet) during topic modeling.

Results

Data Model for Patient Secure Messages

The data model for patient secure messages has a 3-level hierarchical structure consisting of 3 macroconcepts, 28 mesoconcepts, and 85 microconcepts. Textbox 1 provides a partial illustration of the data model (refer to Multimedia Appendix 1 for the full data model). The 3 macroconcepts in the data model are the foundation and base macroconcept, clinical macroconcept, and financial macroconcept. Foundation and base concepts are the basic infrastructure of the health care system concepts on which the rest of the specifications are built. The mesoconcepts of the foundation and base macroconcept include patient, practitioner, related person, organization, health care service, device, appointment, encounter, and document reference. Clinical macroconcepts refer to core clinical components, including allergy intolerance, adverse event, body structure, specimen, condition, procedure, family member history, observation, laboratory test, imaging, medication, immunization, care plan, care team, referral, and risk. The financial macroconcepts cover all finance-related issues such as coverage eligibility, claim payment, account, and explanation of benefits. All the mesoconcepts have been further categorized into microconcepts as attributes, and each mesoconcept has an *unspecified* microconcept. For example, as shown in Textbox 1, the mesoconcept *patient* is categorized as *unspecified*, *privacy*, *lifestyle*, and *diet*. The mesoconcept *coverage-eligibility* has 5 microconcepts, including *unspecified*, *percentage*, *insurance ID*, *benefit category*, *insurer*, and *insurance*.

XSL•FO
RenderX

Textbox 1. A partial illustration of the data model for patient secure messages.

Health Care System Concepts and Their Definitions

Foundation and base concepts

- Patient—Demographic and other administrative information about an individual receiving health care services
  - Unspecified
  - Privacy
  - Lifestyle
  - Diet
- Appointment—A booking of a health care event between patients and practitioners (or other related persons or devices) on a specific date at a specific time
  - Unspecified
  - Status
  - Type
  - Reason

Clinical concepts

- Specimen—A sample taken from a biological entity for laboratory analysis
  - Unspecified
  - Name
- Laboratory test—Tests (eg, clinical, hematological, or microbiology tests) performed on patients and groups of patients and the results derived from the tests
  - Unspecified
  - Name
  - Result

Financial concepts

- Coverage eligibility—Information on patients, insurers, insurance, coverage, plan details, reimbursement, and payment for health care services
  - Unspecified
  - Percentage
  - Insurance ID
  - Benefit category
  - Insurer
  - Insurance
- Explanation of benefits—Information on claim details and adjudication details from the processing of claims
  - Unspecified

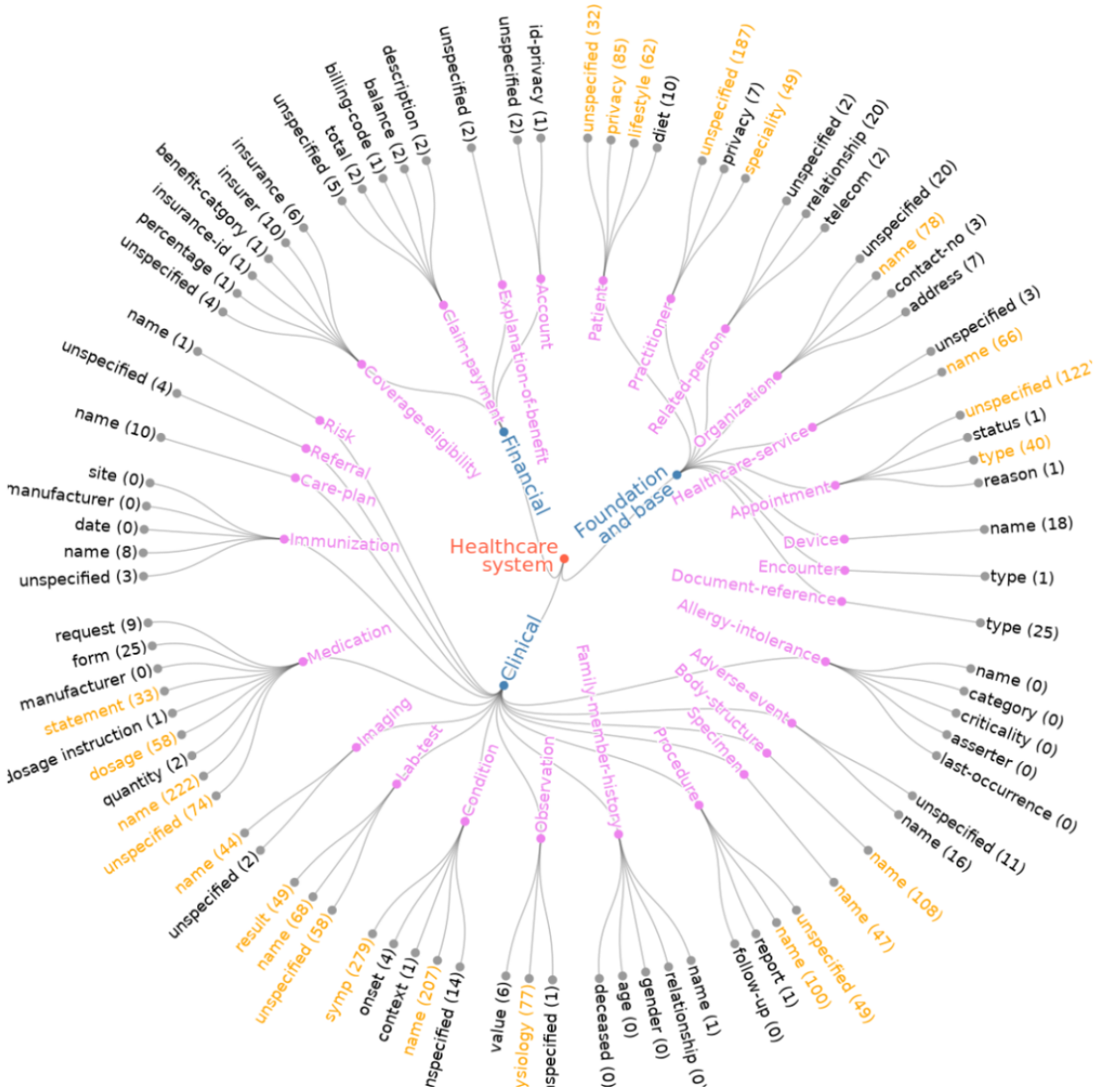
Annotated Corpus of Patient Secure Messages

We annotated 1200 sentences of patient secure messages based on the data model. Figure 2 illustrates the frequency of hierarchical health concepts in the annotated corpus. The concepts in blue, purple, and black represent macroconcepts, mesoconcepts, and microconcepts, respectively. The concepts in orange illustrate the most frequent microconcepts along with their occurrences in the annotated text. Foundation and base macroconcepts make up 33.99% (841/2474) of the annotated corpus, clinical macroconcepts make up 64.38% (1593/2474) of the annotated corpus, and financial macroconcepts make up 1.61% (40/2474) of the annotated corpus, respectively. Patients shared some information about insurance, coverage, and payments in the secure messages. Among the 28 mesoconcepts, the most discussed were condition (505/2474, 20.41%), medication (424/2474, 17.13%), practitioner (243/2474, 9.82%), patient (189/2474, 7.63%), laboratory test (175/2474, 7.07%), appointment (164/2474, 6.62%), procedure (150/2474, 6.06%), and organization (108/2474, 4.36%). The most frequently used

XSL•FO
RenderX

microconcepts were various condition names (eg, fatigue), medication names (eg, prednisone), and appointment types (eg, patient visit).

Figure 2. Radial tree to illustrate hierarchical health concepts in the annotated corpus.

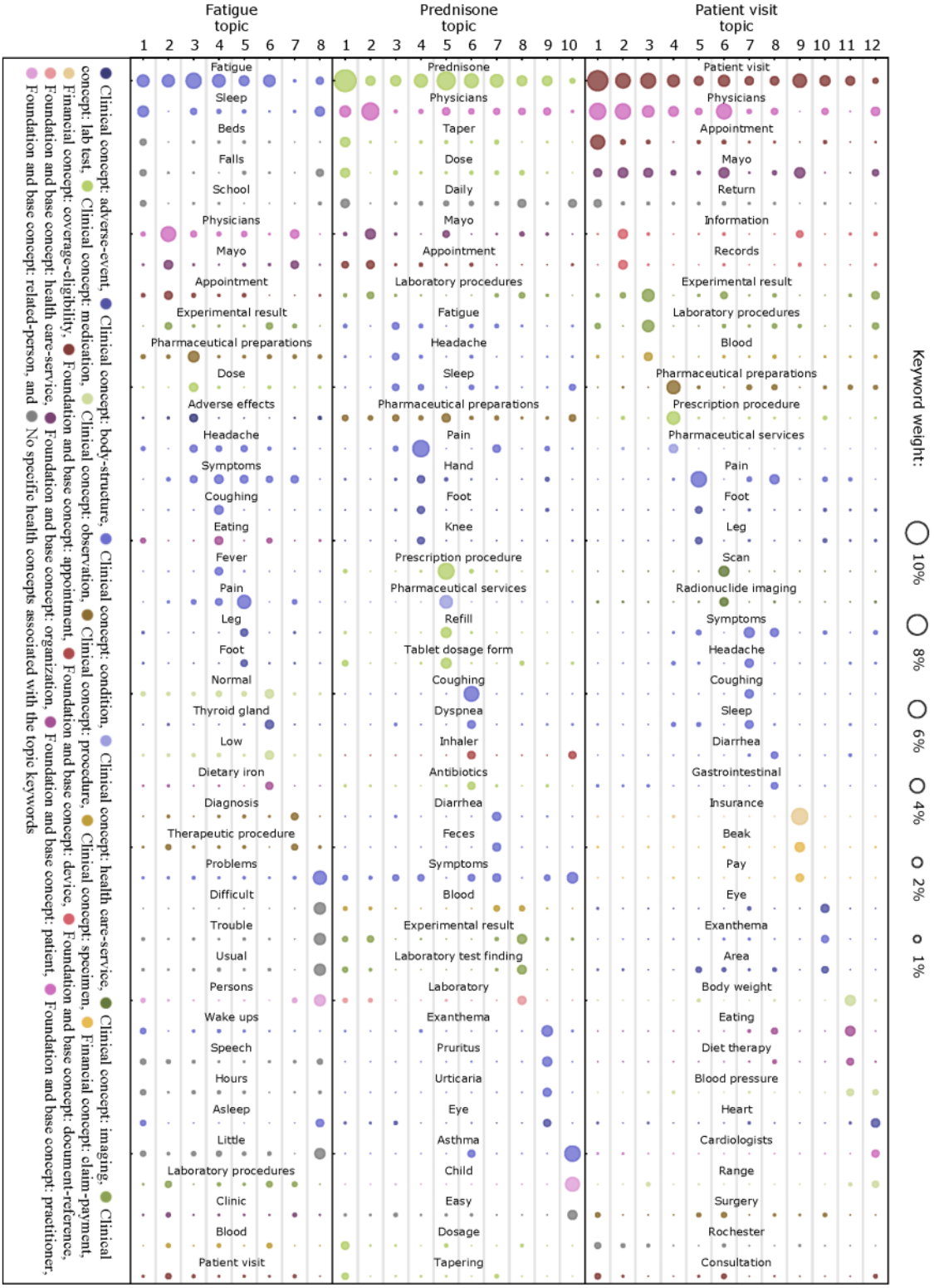


Topics of Patient Secure Messages

After identifying the frequently discussed health concepts, we used a topic modeling technique to discover the latent topics in terms of a group of keywords for patient secure messages mentioning fatigue, prednisone, and patient visit. As shown in Figure 3, we highlight 8, 10, and 12 meaningful topics of patient secure messages related to fatigue, prednisone, and patient visit, respectively. We use a color scheme to represent the top health microconcepts in the topics of patient secure messages associated with fatigue, prednisone, and patient visit. More topic keywords and details can be found in Multimedia Appendix 4. Among the 120 top keywords, 107 (89.2%) were found to be health concepts in the data model. The concepts that were not mapped include general English words such as *hours*, *trouble*, and *difficult*.

XSL•FO
RenderX

Figure 3. The distribution of 40 keywords representing different topic themes (topic 1-12) for the microconcepts fatigue, prednisone, and patient visit. The color indicates different health concepts linked to the topic keywords.



Fatigue

Fatigue is a common clinical condition discussed by patients in the patient secure messages. Topic modeling on patient secure messages targeting fatigue showed that most of the patient discussions were centered around sleepiness, adverse drug effects, and relevant symptoms and conditions. More specifically, the keywords *sleep*, *beds*, *fall*, *asleep*, *little*, and *wake ups* frequently appeared in these patient secure messages, which highlighted the strong association between fatigue and sleepiness. The topic words *pharmaceutical preparations*, *dose*, *adverse effects*, and *headache* revealed that fatigue might be an adverse effect related to some drugs. Other relevant symptoms and conditions such as *headache*, *coughing*, *fever*, and *pain* were also found in patient secure messages discussing fatigue. After tracing back to the original patient secure messages, we

XSL•FO
RenderX

also found mentions of *low*, *thyroid gland*, and *dietary iron*. This finding is consistent with studies reporting that fatigue is one of the most common signs of an underactive thyroid (hypothyroidism) [37] and iron deficiency [38].

Prednisone

Prednisone is a corticosteroid drug that controls inflammation by suppressing the human immune system. Prednisone appeared frequently in the patient secure messages. The focus of patient secure messages related to prednisone was primarily on drug use, side effects, and disease treatment. The topic words related to prednisone use and dose such as *taper*, *dose*, *daily*, *dosage*, and *tapering* were mentioned most often. Prednisone has been used to treat a variety of medical conditions, as shown by the keywords such as *cough*, *dyspnea*, *asthma*, *pruritus*, and *urticaria* [39-42]. In addition, the patients reported side effects such as *fatigue*, *headache*, *sleep* and *pain* after taking prednisone [43].

Patient Visit

Patient visit is a type (attribute) of patient *appointment* under the *foundation and base* concept. The topics of patient secure messages related to patient visits revealed various potential reasons for patient visits by making appointments. For example, many patients requested appointments to visit the Mayo Clinic by mentioning their health information and medical records (*mayo*, *information*, and *records*), laboratory tests and test results (*experimental result*, *laboratory procedure*, *blood*, *scan*, and *radionuclide imaging*), and medications (*pharmaceutical preparation*, *prescription procedure*, and *pharmaceutical services*). The keywords *pain*, *symptoms*, *headache*, *coughing*, *diarrhea*, *exanthema*, *heart*, and *cardiologists* suggest the purpose of their visit to the Mayo Clinic. In addition, some patients made appointments for financial issues indicated by the topic words *insurance* and *pay*.

Discussion

Principal Findings

Patient secure messages contain valuable information about the quality of health care delivery systems, drug efficacy and safety signals, and other pain points of patient health management. The data model and annotated corpus enable us to extract information from patient secure messages for systematic content analysis and to resolve the challenges of data analysis and integration by standardizing the unstructured data with the structured data system. In future, this data model can be used and extended to model patient narratives from social media platforms [44] to analyze their content. Our future aim is to leverage this data model and annotated corpus for developing machine learning models and NLP-enabled parsing tools for automatically parsing patient narratives to advance clinical research and practice [45,46].

During annotation, we faced some challenges because of the heterogeneous and superficial nature of the language used in the messages, which often deviated from formal English grammar, spelling, and punctuation rules. These challenges can often lower the quality of the data and make them less accessible to automated processing by a system. We discussed and designed some solutions to overcome these challenges. The challenges and their solutions are discussed in detail in Multimedia Appendix 2. During the comparison between the 2 sets of the first set of annotation, the level of agreement between the annotators was consistent for a few subconcepts, such as *appointment*, *diagnostic report*, *immunization*, *medication*, and *practitioner*. In contrast, it was very unsatisfactory for subconcepts such as *specimen*, *related-person*, *document-reference*, *eligibility*, and *health care-service*. Although the 2 annotators achieved a satisfactory IAA score later, annotation bias likely exists in the annotated corpus.

After analyzing the annotated corpus and identifying important health concepts in patient secure messages, we further used topic modeling to automatically uncover hidden topics of patient secure messages mentioning health concepts [47]. In this way, we were able to evaluate the data model and understand the focus of patient discussion and concerns in the patient secure messages. For example, prednisone, because it is a corticosteroid drug that controls inflammation, was a highly discussed topic in the patient secure messages. The patient-provider discussion on prednisone was primarily centered around medication use, side effects, and disease treatment [39-43]. These findings offer useful information for shared clinical decision making and patient-centered care. LDA exploits statistical inference to identify latent topics using a bag-of-words model and term frequency. Therefore, LDA mainly discovers topics with high frequency and dominant terms and pays little attention to rare, yet meaningful, topics from patient secure messages. In addition, hypermeter tuning in LDA can be more art than science.

Our study is the first to develop a data model based on HL7 FHIR to understand and analyze the content of patient secure messages. This study has several limitations. For annotation, the F1 scores were 0.62 (macro mean) and 0.74 (micro mean). This indicates that the task of annotation is a difficult one because we need to assign 3-level hierarchical health concepts to an identified health entity, and there are 85 microconcepts to be selected in the data model. Our annotators not only need to assign a category to these potential entities but also to identify their boundaries. If the boundaries are defined differently by different annotators, it is considered *notconsistent*. The language used in patient secure messages to describe medical concepts is casual, colloquial, and ambiguous. We revised our annotation guideline 5 times and trained our annotators 4 times. We believe that the F1 score could be improved, given more resources for annotation.

All the messages analyzed in this study were sent by patients. We did not analyze the messages generated by the clinicians who read and replied to these patient secure messages. The Mayo Clinic at Rochester, Minnesota, is a large nonprofit academic medical center that provides comprehensive patient care in the United States. We acknowledge that the patients and medical practices at the Mayo Clinic are not necessarily a representative cross-section of all patients and medical practices in the country. It is also unknown how the data model can be applied to other hospitals in other countries, given different models of care and patient-clinician relationships. For example, we are aware that in China, there are no patient portals, and patients communicate with their providers through other means.

XSL•FO
RenderX

Thus, we share our model, annotation guideline, and other materials with the broader scientific community and welcome all sorts of collaboration or partnership.

Finally, we acknowledge that evaluation and validation are key and challenging in this case. Strictly speaking, the data model can only be validated for the claimed utilities in real-world implementations. The topic modeling analysis we conducted was only an initial and weak evaluation. We expect to conduct a much more rigorous evaluation and validation by building collaborations and partnerships with domestic and international researchers in the field and by moving forward with building the NLP systems.

In the next stage of our study, we aim to curate a larger annotated corpus and use it for training and testing machine learning models for automated triaging. The manually annotated corpus is reusable for future NLP research to save manual effort and cost, although there might be some challenges related to its reuse because of data privacy and confidentiality challenges. This corpus is also generated based on patient secure messages; therefore, there always will remain a question of its usability in different domains of medical research.

Conclusions

A patient portal as a tethered EHR system enables patients to access their medical records, seek support, and share their opinions with their caregivers through secure messaging between their clinical visits. The large volume of secure messages opens new opportunities and challenges for understanding patient concerns and information integration into EHRs to improve patient-centered care. This study is a novel attempt to identify the content of patient secure messages based on the foundation and base, clinical, and financial concepts of HL7 FHIR standards. The data model and annotated corpus enable us to meet the challenges of analyzing and understanding unstructured health information from patient secure messages along with the topic modeling technique to discover the hidden topics on interesting health concepts in patient secure messages.

The data model could be potentially used for automatically identifying and analyzing other types of patient narratives such as those in various social media and patient forums. In the future, we plan to develop a machine learning and NLP solution to enable automatic triaging solutions to reduce the workload of clinicians and perform more granular content analysis to understand patients' needs and improve patient-centered care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data model for patient secure messages based on Health Level-7 Fast Healthcare Interoperability Resources.
[XLSX File (Microsoft Excel File), 13 KB - jmir_v23i7e26770_app1.xlsx]

Multimedia Appendix 2
Description of methodology in detail.
[PDF File (Adobe PDF File), 285 KB - jmir_v23i7e26770_app2.pdf]

Multimedia Appendix 3
Annotation guideline.
[PDF File (Adobe PDF File), 343 KB - jmir_v23i7e26770_app3.pdf]

Multimedia Appendix 4
Results of topic modeling.
[XLSX File (Microsoft Excel File), 33 KB - jmir_v23i7e26770_app4.xlsx]

References

1. Casalino LP. The medicare access and CHIP reauthorization act and the corporate transformation of American medicine. Health Aff 2017 May 1;36(5):865-869. [doi: 10.1377/hlthaff.2016.1536] [Medline: 28461353]
2. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
3. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
4. Plastiras P, O'Sullivan D, Weller P. An Ontology-driven Information Model for Interoperability of Personal and Electronic Health Records. ResearchGate. 2014 Apr. URL: https://www.researchgate.net/publication/313512132_An_Ontology-Driven_Information_Model_for_Interoperability_of_Personal_and_Electronic_Health_Records [accessed 2019-01-20]

XSL•FO
RenderX

5. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
6. Medicare and Medicaid Programs; Electronic Health Record Incentive Program-stage 2. Federal Register. 2012. URL: https://www.federalregister.gov/documents/2012/09/04/2012-21050/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-2 [accessed 2020-02-13]
7. Henry J, Barker W, Kachay L. Electronic Capabilities for Patient Engagement among U.S. Non-Federal Acute Care Hospitals: 2013-2017. Office of the National Coordinator for Health Information. 2019 Apr. URL: https://www.healthit.gov/sites/default/files/page/2019-04/AHApatientengagement.pdf [accessed 2020-02-13]
8. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
9. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 3;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
10. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011 May 26;13(2):e41 [FREE Full text] [doi: 10.2196/jmir.1702] [Medline: 21616784]
11. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
12. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
13. Cronin R, Davis S, Shenson J, Chen Q, Rosenbloom S, Jackson G. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]
14. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]
15. Masterman M, Cronin R, Davis S, Shenson J, Jackson G. Adoption of Secure Messaging in a Patient Portal across Pediatric Specialties. AMIA Annu Symp Proc 2016;2016:1930-1939 [FREE Full text] [Medline: 28269952]
16. HL7 FHIR Release 4. Health Level Seven International. 2019. URL: https://www.hl7.org/fhir/ [accessed 2019-11-04]
17. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
18. Sulieman L, Robinson JR, Jackson GP. Automating the classification of complexity of medical decision-making in patient-provider messaging in a patient portal. J Surg Res 2020 Nov;255:224-232 [FREE Full text] [doi: 10.1016/j.jss.2020.05.039] [Medline: 32570124]
19. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. Int J Med Inform 2017 Sep;105:110-120 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.06.004] [Medline: 28750904]
20. About HL7. Health Level Seven International. 2019. URL: https://www.hl7.org/about/ [accessed 2020-01-20]
21. What Is FHIR? Office of the National Coordinator for Health Information Technology. 2019. URL: https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf [accessed 2020-01-20]
22. Saripalle R. Fast health interoperability resources (FHIR): current status in the healthcare system. Int J E-Health Med Commun 2019 Jan;10(1):76-93. [doi: 10.4018/IJEHMC.2019010105]
23. Hong N, Wen A, Shen F, Sohn S, Liu S, Liu H, et al. Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. AMIA Jt Summits Transl Sci Proc 2018;2017:74-83 [FREE Full text] [Medline: 29888045]
24. Patient Online Services. Mayo Clinic. 2010. URL: https://onlineservices.mayoclinic.org/content/staticpatient/showpage/patientonline [accessed 2018-10-02]
25. Stubbs A. MAE and MAI: Lightweight Annotation and Adjudication Tools. In: Proceedings of the 5th Linguistic Annotation Workshop. 2011 Jun Presented at: LAW V'11; June 23-24, 2011; Portland, Oregon, USA p. 129-133.
26. Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002 Jul 6 Presented at: ACL'02; July 6-12, 2002; Philadelphia, PA, USA p. 168-175.
27. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
28. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 1;3:993-1022.

XSL•FO
RenderX

29. McCallum AK. Mallet: A machine learning for language toolkit. Mallet. 2002. URL: http://mallet.cs.umass.edu [accessed 2019-05-15]
30. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 Feb Presented at: WSDM'15; January 31-February 6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
31. Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
32. Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform 2012 Oct;45(5):885-892 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.008] [Medline: 22554702]
33. Blackman NJ, Koval JJ. Interval estimation for Cohen's kappa as a measure of agreement. Statist Med 2000 Mar 15;19(5):723-741. [doi: 10.1002/(sici)1097-0258(20000315)19:5<723::aid-sim379>3.0.co;2-a] [Medline: 10700742]
34. How FHIR Fits into an EHR. Health Level Seven International. 2019. URL: https://www.hl7.org/fhir/ehr-fm.html [accessed 2019-12-10]
35. Huang M, ElTayeby O, Zolnoori M, Yao L. Public opinions toward diseases: infodemiological study on news media data. J Med Internet Res 2018 May 8;20(5):e10047 [FREE Full text] [doi: 10.2196/10047] [Medline: 29739741]
36. Huang M, Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Yao L. Technological innovations in disease management: text mining US patent data from 1995 to 2017. J Med Internet Res 2019 Apr 30;21(4):e13316 [FREE Full text] [doi: 10.2196/13316] [Medline: 31038462]
37. Chaker L, Bianco AC, Jonklaas J, Peeters RP. Hypothyroidism. Lancet 2017 Sep 23;390(10101):1550-1562 [FREE Full text] [doi: 10.1016/S0140-6736(17)30703-1] [Medline: 28336049]
38. Camaschella C. Iron-deficiency anemia. N Engl J Med 2015 May 7;372(19):1832-1843. [doi: 10.1056/NEJMra1401038] [Medline: 25946282]
39. Thompson WH, Nielson CP, Carvalho P, Charan NB, Crowley JJ. Controlled trial of oral prednisone in outpatients with acute COPD exacerbation. Am J Respir Crit Care Med 1996 Aug;154(2):407-412. [doi: 10.1164/ajrccm.154.2.8756814] [Medline: 8756814]
40. Schuh S, Reisman J, Alshehri M, Dupuis A, Corey M, Arseneault R, et al. A comparison of inhaled fluticasone and oral prednisone for children with severe acute asthma. N Engl J Med 2000 Sep 7;343(10):689-694. [doi: 10.1056/NEJM200009073431003] [Medline: 10974132]
41. Taranta A, Mark H, Haas RC, Cooper NS. Nodular panniculitis after massive prednisone therapy. Am J Med 1958 Jul;25(1):52-61. [doi: 10.1016/0002-9343(58)90198-0]
42. Pollack CV, Romano TJ. Outpatient management of acute urticaria: the role of prednisone. Ann Emerg Med 1995 Nov;26(5):547-551. [doi: 10.1016/s0196-0644(95)70002-1]
43. Prednisone. Mayo Clinic. 2020. URL: https://www.mayoclinic.org/drugs-supplements/prednisone-oral-route/side-effects/drg-20075269?p=1 [accessed 2019-08-10]
44. Ru B, Harris K, Yao L. A Content Analysis of Patient-Reported Medication Outcomes on Social Media. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). 2015 Nov Presented at: ICDMW'15; November 14-17, 2015; Atlantic City, NJ, USA. [doi: 10.1109/icdmw.2015.150]
45. Soni S, Gudala M, Wang D, Roberts K. Using FHIR to construct a corpus of clinical questions annotated with logical forms and answers. AMIA Annu Symp Proc 2019;2019:1207-1215 [FREE Full text] [Medline: 32308918]
46. Zolnoori M, Fung KW, Patrick TB, Fontelo P, Kharrazi H, Faiola A, et al. A systematic approach for developing a corpus of patient reported adverse drug events: a case study for SSRI and SNRI medications. J Biomed Inform 2019 Feb;90:103091 [FREE Full text] [doi: 10.1016/j.jbi.2018.12.005] [Medline: 30611893]
47. Zolnoori M, Huang M, Patten CA, Balls-Berry JE, Goudarzvand S, Brockman TA, et al. Mining news media for understanding public health concerns. J Clin Transl Sci 2019 Oct 23;5(1):e1 [FREE Full text] [doi: 10.1017/cts.2019.434] [Medline: 33948233]

Abbreviations

EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level-7
IAA: interannotator agreement
LDA: latent Dirichlet allocation
NLP: natural language processing

XSL•FO
RenderX

Edited by R Kukafka; submitted 06.01.21; peer-reviewed by X Yang, M Rigby; comments to author 16.03.21; revised version received 10.05.21; accepted 24.05.21; published 30.07.21.

Please cite as:
De A, Huang M, Feng T, Yue X, Yao L
Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)–Based Data Model: Development and Topic Modeling Study
J Med Internet Res 2021;23(7):e26770
URL: https://www.jmir.org/2021/7/e26770
doi:10.2196/26770
PMID:34328444

©Amrita De, Ming Huang, Tinghao Feng, Xiaomeng Yue, Lixia Yao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation

András Lánczky[1,2], MSc; Balázs Győrffy[1,2], MD, PhD, DSC

[1]Department of Bioinformatics, Semmelweis University, Budapest, Hungary

[2]TTK Lendület Cancer Biomarker Research Group, Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

Corresponding Author:
Balázs Győrffy, MD, PhD, DSC
Department of Bioinformatics
Semmelweis University
Tűzoltó utca 7-9
Budapest, 1094
Hungary
Phone: 36 305142822
Email: gyorffy.balazs@med.semmelweis-univ.hu

Abstract

Background: Survival analysis is a cornerstone of medical research, enabling the assessment of clinical outcomes for disease progression and treatment efficiency. Despite its central importance, no commonly used spreadsheet software can handle survival analysis and there is no web server available for its computation.

Objective: Here, we introduce a web-based tool capable of performing univariate and multivariate Cox proportional hazards survival analysis using data generated by genomic, transcriptomic, proteomic, or metabolomic studies.

Methods: We implemented different methods to establish cut-off values for the trichotomization or dichotomization of continuous data. The false discovery rate is computed to correct for multiple hypothesis testing. A multivariate analysis option enables comparing omics data with clinical variables.

Results: We established a registration-free web-based survival analysis tool capable of performing univariate and multivariate survival analysis using any custom-generated data.

Conclusions: This tool fills a gap and will be an invaluable contribution to basic medical and clinical research.

(J Med Internet Res 2021;23(7):e27633) doi:10.2196/27633

KEYWORDS

Kaplan-Meier plot; internet; Cox regression; follow-up; multivariate analysis; survival

Introduction

Bioinformatic programs include databases, algorithms, services, and software tools. These not only span a wide range of utility but have also gained increased value in scientific research in recent years; approximately 80% of papers published in biology and 60% of papers published in medicine report the use of at least one bioinformatic tool [1]. We recently analyzed the landscape of web-based bioinformatic services and uncovered 3649 such publications since 1994, 69% of which are actively maintained [2]. The leading advantages of browser-based bioinformatic programs include unrestricted availability, the lack of need for the installation of specific software packages, optimized allocation of computational resources, the possibility of constant updates, instant access to the latest versions, and the opportunity to enable real-time validation of previous analysis results. A subcohort of these tools enables certain analyses with user-provided data. A few representative examples used in these tools in medical research include an online calculator for receiver operator characteristics [3], a tool to determine optimal cut-off values for clinical tests [4], and a sample size calculator for randomized clinical trials [5].

The assessment of survival following the onset of a disease or of a treatment is a fundamental analysis in medical research. In an optimal scenario, the differential survival of two cohorts can be compared by employing a simple Mann-Whitney test. However, survival times do not follow a normal distribution and it is common for numerous subjects to lack associated event data at the end of follow-up. Kaplan and Meier [6] proposed a simple and elegant solution to these issues by including all cases

XSL•FO
RenderX

regardless of endpoint status in the analysis. The basic concept of Kaplan-Meier survival analysis is to assign a "censored" status to incomplete observations at the end of the follow-up time. In other words, there are two inputs for each case: the length of the follow-up time and a binary classifier designating the case as one with an event or one that is censored. Then, starting from 100, at each event, the survival line drops in proportion to the number of samples remaining within the investigated cohort. If there are multiple survival curves, the statistical difference between these is most commonly computed by employing the Cox proportional-hazards regression model [7].

Despite its widespread use, there is no online tool available for survival analysis. Therefore, it is necessary to acquire specialized software packages, as none of the general office packages (OpenOffice, LibreOffice, MS Office) is suitable for analyzing follow-up data. We previously established an online platform capable of linking survival outcome in various cancer types to mRNA [8] and microRNA [9] expression alterations. Here, we aimed to establish a freely available, easy-to-use online platform capable of performing survival analysis and constructing a Kaplan-Meier plot with any type of user-uploaded custom data containing any type of genomic or clinical information.

Methods

Setup of the Web Platform

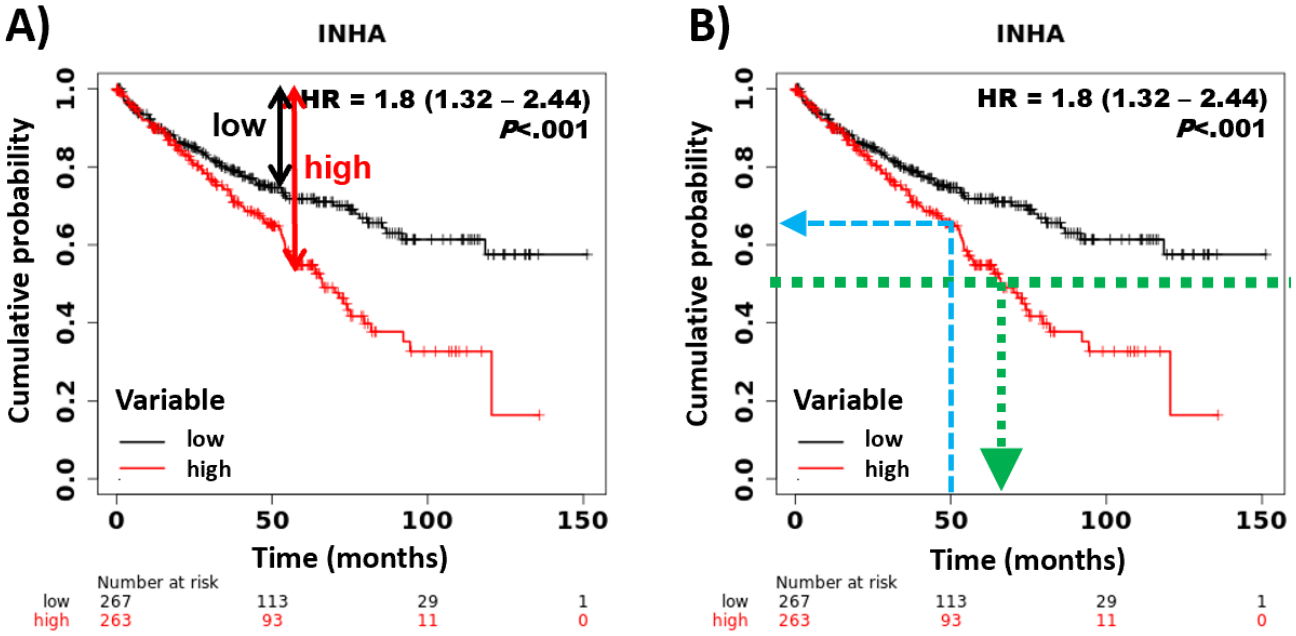
The website is built on an Apache 2.4 web server and hosted by a Linux-based server machine. The user interface is written in PHP 7 and JavaScript using JQuery. The backend side is written in PHP 7 and R, and the repository layer is built on the PostgreSQL 12 database. The database temporarily contains the uploaded data and generated results. The analysis platform is accessible via any standard browser (Firefox, Edge, Chrome, Safari).

Survival Analysis

Multiple R packages are used for the statistical computations and for generating the output graphs. The *survival* package [10] is used for univariate Kaplan-Meier analysis and the multivariate analysis. The survival curve and the beeswarm plot are generated by the *survplot* [11] and *beeswarm* [12] packages, respectively. The *XML* and *rjson* R packages are used to load the configuration files, the *RODBC* package is used to communicate with the database, and the *ggplot2* package [13] is used to visualize the results.

When comparing two cohorts, the significance is computed using the Cox-Mantel (log rank) test [7]. The difference between the cohorts is numerically characterized by the hazard rate (HR), which is based on the differential descent rate of the two cohorts (Figure 1A). Of note, since the hazard rate is by definition a comparison to the baseline, a relative two-fold drop in one cohort is equal to a half-fold drop in the other cohorts. Basically, depending on the context, an HR of 2 equals an HR of 0.5. As it is easier to understand an HR value above 1 in most cases, we implemented an option to invert all HR values below 1.

Figure 1. Kaplan-Meier curves showing main concepts used in survival analysis, including the (A) hazard rate (high/low) and (B) median survival. The green arrow shows the visually determined median survival and the blue arrow shows the survival probability at 50 months.



The generated results also include the median survival time, which is the time at which the probability of 0.5 is reached in one of the cohorts. The median time can also be determined visually by drawing a vertical line from the selected probability to the X axis. Of note, performing the steps backward can determine the cumulative probability of survival at a given time point (Figure 1B).

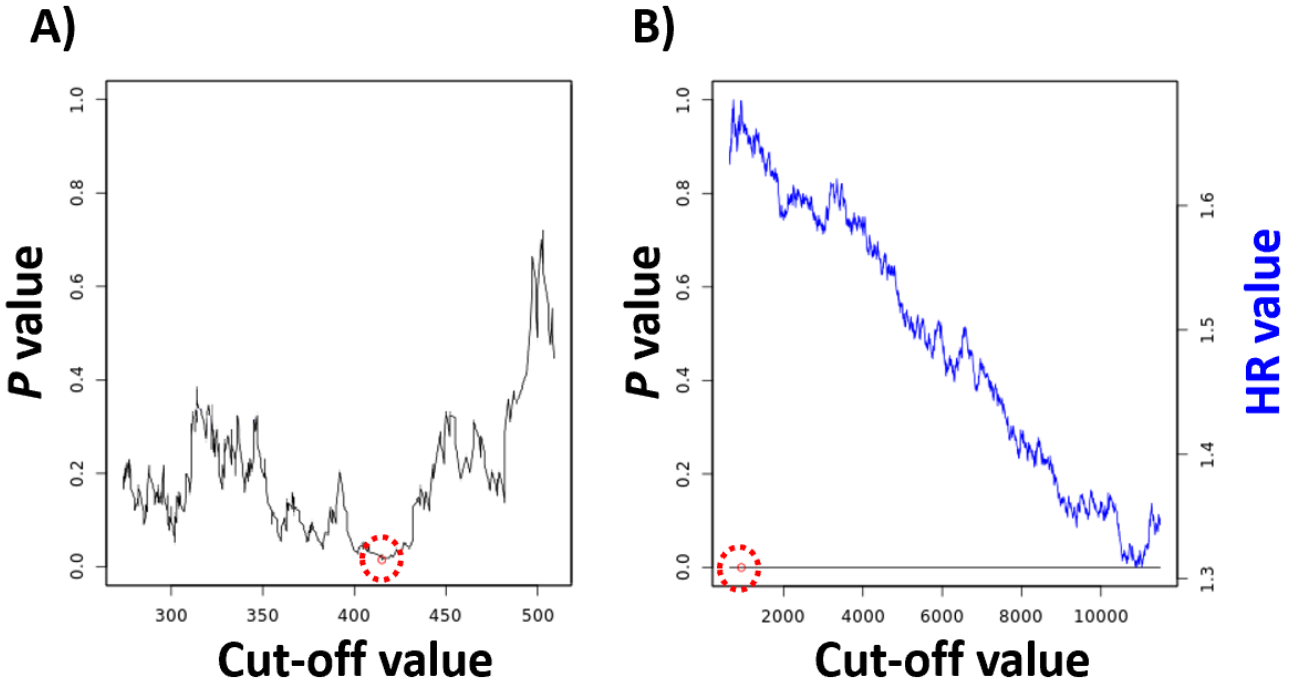
XSL•FO
RenderX

Assigning the Samples into Two Cohorts

To enable visualization in the Kaplan-Meier plot, it is necessary to establish a cut-off value and assign the samples to one of two cohorts. We implemented three different options for this task: using a predefined quantile (including the median, upper, and lower quartiles), trichotomizing the data (eg, assign the data into three cohorts and then omit the middle cohort), and using the best available cut-off value.

To find the best cutoff, we iterate over the input variable values from the lower quartile to the upper quartile and compute the Cox regression [7] for each setting. The most significant cut-off value is used as the best cutoff to separate the input data into two groups. We implemented a simple visual representation of this analysis by showing the achieved *P* values in relation to the used cut-off values (Figure 2A). In case the generated cut-off values are ambiguous (eg, multiple cut-off values deliver very low *P* values), the cut-off value corresponding to the highest HR is used (see Figure 2B).

Figure 2. A cut-off plot can be used to visualize the correlation between the used cut-off values and the achieved *P* values (A) and hazard rate (HR) (B). The red circle identifies the best cutoff. The computation of false discovery rate across all *P* values provides correction for multiple hypothesis testing.



Quality Control

During the computation of multiple cut-off values, multiple hypotheses are generated. Therefore, the false discovery rate (FDR) is computed by default in this setting using the Benjamini-Hochberg method [14] to correct for multiple hypothesis testing. The FDR results are normally shown in the "Results" page.

A requirement for Cox regression is that the hazard is independent of time. To fulfill this requirement, the censoring should be independent of the prognosis, samples entering at different time points in the analysis should have the same prognosis, and the time should be measured as a continuous variable (not in bins). We employed the coxph function of the *survival* package [10] for performing the proportional hazard assumption test.

In some cases, one might want to compare clinical and genomic variables. To enable this, clinical data can be selected not only as filters but also as variables to be included in the multivariate analysis. In these analyses, the "Results" page displays the *P* values and HR values for each variable included in the multivariate analysis in a table format.

Using Multiple Variables

We implemented multiple options to simultaneously use and combine multiple variables. Each of these settings uses the original variable values as input and basic mathematical functions to calculate the new joint values.

The simplest option is to select multiple variables and then use each variable separately. In this case, the same analysis is performed for each selected marker using the exact same filtering settings. This option is identical to running the analysis for each variable consecutively.

In the second feature, one can use the mean expression of a panel of variables; in this case, any variable can be inverted and a weight can be added to each. Using the mean expression of a set of genes can be termed a "signature analysis," as the expression of each included variable will influence the value of the final "composite variable." This feature can also be used to validate previously published gene expression signatures utilizing a preselected panel of genes.

A third option is utilization of the ratio of two genes; in this case, one variable is used as the numerator, the other variable is used as the denominator, and a new value is computed for

XSL•FO
RenderX

each sample. This setting is useful when one wants to compare the expression values to a reference gene such as *GAPDH*.

The fourth option enables the stratification of all patients based on the median expression level of a selected variable and then use another variable in the high or low cohort only. This enables the investigation of a selected variable in an already stratified cohort and ultimately the setup of a decision tree–like classification for the investigated cohort.

In each of the settings where multiple variables are combined, a new value based on the equation is generated for each sample, which is then used when performing the survival analysis, including the cut-off selection. Of note, one might want to directly compare two or more selected continuous variables to each other. For this purpose, we implemented an option to compute Spearman and Pearson correlation coefficients between the variables using the *cor.test* function from the basic R distribution.

Results

We established an online survival analysis platform that grabs a user-generated tab-separated or semicolon-separated file as input. The table headers can include case-insensitive letters of the English alphabet, numbers, spaces, underscores, colons, round brackets, and exclamation marks as characters. The content within the table cells can be numeric or text values. Some columns can be used as filters and a maximum of three filters are allowed. Table 1 provides a quick guide for the setup of an input file. The file can be a comma-separated value or a tab-separated table, and different types of data are allowed in each column. Table 2 shows a sample input file using this guide; the maximal dimensions of the table are 100 columns and 8000 rows. Note that a gene can be in the form of text when used as a group. Using this data table, the system is capable of performing univariate and multivariate survival analysis by using one or multiple variables and clinical data. In addition to drawing a Kaplan-Meier plot, the *P* values and HR values with 95% CIs are also computed. A separate plot visualizes the correlation between the *P* values and HR values and the employed cut-off values. Median survival values are computed for cohorts reaching a cumulative probability of 0.5, and upper-quartile survival is computed for the remaining cases. Of note, when performing multivariate analysis, only patient samples for which all variables of interest are concurrently available can be included. The platform includes multiple quality-control steps, including validation of the proportional hazard assumption and computation of the FDR for cases where multiple analyses are run simultaneously. The web service is freely available without requiring registration [15].

Table 1. Quick start guide for setting up an input file.

Header name	Sample ID	Survival time	Survival event	Filter	Gene
Automatically recognized	Yes	Yes	Yes	Yes	No
Maximal number of different values	No limit	No limit	2 (0 or 1)	10	No limit
Can be text	Yes	No	No	Yes	No
Can be binary	No	No	Yes	Yes	Yes
Can be continuous		Yes	No	No	Yes

Table 2. Sample input file.

Sample ID	Survival time	Survival event	Filter_A	Filter_B	Filter_C	Gene_1	ABC123	DE45
Sample 1	95	1	2	2	3	1441	4474	1.13
Sample 2	66	0	3	3	3	3064	421	2.395
Sample 3	70	0	3	1	1	2529	2974	1.363
Sample 4	26	1	3	1	3	19	3346	4.818
Sample 5	13	0	1	2	3	3573	1244	2.058
Sample 6	67	0	2	3	2	2977	962	4.431
Sample 7	96	1	3	3	3	2777	4367	2.015
Sample 8	67	0	3	3	1	4606	4190	1.05
Sample 9	95	1	3	1	2	1209	3930	1.980
Sample 10	1	1	2	3	2	1894	4897	4.073

Discussion

Currently, genomics, transcriptomics, proteomics, and metabolomics enable the simultaneous investigation of multiple markers related to patient prognosis in experimental and clinical studies. Multiple online tools make survival analysis possible using previously published datasets such as those employing data from The Cancer Genome Atlas [9,16]. Despite the almost ubiquitous use of Cox regression to correlate different marker levels to prognosis, there is no available software to perform survival analysis for user-generated custom datasets. We

XSL•FO
RenderX

established a wide-ranging online tool capable of performing Cox regression and constructing Kaplan-Meier plots for user-generated data. A comprehensive and practical review of the Kaplan-Meier curves has been published previously [17].

A major advantage of our platform is the inclusion of multiple choices to select a cut-off value to be used in the analysis. To generate a Kaplan-Meier plot, one must first determine a cutoff; a convenient and widespread option for this task is the median expression value [18,19]. However, the cutoff should be based on the intention of the study. In most medical studies, there is no biological reason that a certain predetermined quantile cutoff should discriminate two cohorts [20]. When a researcher aims to uncover any potential correlation between a variable and outcome, then all possible cut-off values can be checked. Of course, in such cases, the chance of false-positive results also increases; therefore, we have implemented the Benjamini-Hochberg method [14] to calculate the FDR to correct for multiple hypothesis testing. Our approach is rather conservative as the different analyses are not truly independent in such a scenario, as only a few samples can switch cohorts between successive analyses. Of note, independent of the used cutoff, a single-variable analysis is almost never sufficient to prove a direct correlation and thus multivariate analysis should not be omitted.

The analysis automatically checks the proportional hazards assumption to evaluate the independence from time. This can also be achieved by a simple visual inspection of the graph: in case there seems to be a significant difference between the two cohorts but the lines cross at multiple time points, then the hazard is clearly not independent of time [21]. Of note, a common question is whether or not crossing at the right end of the plot violates the proportional hazards assumption. In most cases, at the end of the follow-up time, only few patients remain in both cohorts. Thus, because the drop in the line for each event is proportional to all samples remaining in the analysis, even an event for a single patient can result in crossing of the two lines. However, this will not affect the significance of the entire analysis.

When interpreting the results, one has to be aware of some common caveats of survival analysis. First, the *P* value should be interpreted with respect to the sample size. The Cox model is not suitable for small sample sizes (N<40), and in these cases the generalized log-rank method is a better choice [22]. Higher sample numbers will lead to better significance, even in cases where the HR values are lower. A representative example of this bias is the ill-fated FLEX phase III trial [23]. By investigating the effect of cetuximab in patients with advanced nonsmall cell lung cancer, the authors observed a difference in survival of 10.1 months vs 11.3 months in the untreated and treated cohort, respectively. Although this difference was initially considered to be sufficient to gain approval by the US Food and Drug Administration, the European Medicines Agency rejected approval of the drug. Their main problem with the trial was the minimal overall survival benefit of only 12% and that only the exceptionally high sample number (N=1125) helped to reach a minimally significant *P* value of .04 [23].

A second important deception is the proportion of recorded events within a study. As only the actual events contribute to the drops in survival curves, it is not possible to perform a meaningful survival analysis when the number of events is very low. This not only prevents the computation of median (or upper quartile) survival, but the accidental concentration of all events into one of the cohorts can lead to an infinite HR. For example, The Cancer Genome Atlas Network published a breast cancer dataset with approximately 1000 patient samples [24]. The authors had to note that because of the very short follow-up, only 11% of the samples had survival events, which prevented utilization of the dataset for survival analyses [24].

We also have to discuss some limitations of the software. The input file has to be carefully formatted, and a maximum of 100 columns and 8000 rows are allowed. Only full columns are acceptable as variables, a maximum of three filters can be defined, and the survival event can only be coded “0” or “1.” Although these restrictions can make the setup of the analysis challenging, when a correctly formatted table is uploaded, the system can automatically recognize columns representing a survival event or survival time. A second limitation is the exclusive use of the Cox proportional-hazards model to compute significance, and other tests such as the Cochran-Mantel-Haenszel test [25,26] or the Gehan-Breslow-Wilcoxon test [27,28] are not implemented. The reason for our restriction is the almost exclusive use of the Cox test in the current medical literature.

In summary, we established an online survival analysis tool capable of performing univariate and multivariate survival analysis using any custom-generated data. We believe that this registration-free online platform simultaneously integrating multiple different analysis and quality-control options will be a valuable tool for biomedical researchers.

Acknowledgments

The research was financed by the 2018-2.1.17-TET-KR-00001, 2020-1.1.6-JÖVŐ-2021-00013 and 2018-1.3.1-VKE-2018-00032 grants and by the Higher Education Institutional Excellence Program (2020-4.1.1.-TKP2020) of the Ministry for Innovation and Technology (MIT) in Hungary, within the framework of the Bionic thematic program of the Semmelweis University. The authors wish to acknowledge the support of ELIXIR Hungary.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Duck G, Nenadic G, Filannino M, Brass A, Robertson DL, Stevens R. A survey of bioinformatics database and software usage through mining the literature. PLoS One 2016;11(6):e0157989 [FREE Full text] [doi: 10.1371/journal.pone.0157989] [Medline: 27331905]
2. Ősz Á, Pongor LS, Szirmai D, Győrffy B. A snapshot of 3649 web-based services published between 1994 and 2017 shows a decrease in availability after 2 years. Brief Bioinform 2019 May 21;20(3):1004-1010 [FREE Full text] [doi: 10.1093/bib/bbx159] [Medline: 29228189]
3. Goksuluk D, Korkmaz S, Zararsiz G, Karaagaoglu A. easyROC: an interactive web-tool for ROC curve analysis using R language environment. R Journal 2016;8(2):213. [doi: 10.32614/rj-2016-042]
4. Budczies J, Klauschen F, Sinn BV, Győrffy B, Schmitt WD, Darb-Esfahani S, et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One 2012;7(12):e51862 [FREE Full text] [doi: 10.1371/journal.pone.0051862] [Medline: 23251644]
5. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. Chest 2020 Jul;158(1S):S12-S20. [doi: 10.1016/j.chest.2020.03.010] [Medline: 32658647]
6. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958 Jun;53(282):457-481. [doi: 10.2307/2281868]
7. Cox DR. Regression models and life-tables. J R Stat Soc B 2018 Dec 05;34(2):187-202. [doi: 10.1111/j.2517-6161.1972.tb00899.x]
8. Gyorffy B, Lánczky A, Szállási Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer 2012 Apr;19(2):197-208. [doi: 10.1530/ERC-11-0329] [Medline: 22277193]
9. Lánczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, et al. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. Breast Cancer Res Treat 2016 Dec;160(3):439-446. [doi: 10.1007/s10549-016-4013-7] [Medline: 27744485]
10. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer-Verlag; 2000.
11. Survplot. URL: http://www.cbs.dtu.dk/~eklund/survplot [accessed 2021-07-07]
12. BeeSwarm. URL: http://www.cbs.dtu.dk/~eklund/beeswarm/ [accessed 2021-07-07]
13. Wickham H. ggplot2: elegant graphics for data analysis. New York, NY: Springer-Verlag; 2009.
14. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
15. Kaplan-Meier plotter with custom data. URL: https://kmplot.com/custom_data [accessed 2021-07-02]
16. Pak K, Oh S, Goh T, Heo H, Han M, Jeong D, et al. A user-friendly, web-based integrative tool (ESurv) for survival analysis: development and validation study. J Med Internet Res 2020 May 05;22(5):e16084 [FREE Full text] [doi: 10.2196/16084] [Medline: 32369034]
17. Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg 2010 Sep;143(3):331-336 [FREE Full text] [doi: 10.1016/j.otohns.2010.05.007] [Medline: 20723767]
18. Mishra NK, Southekal S, Guda C. Survival analysis of multi-omics data identifies potential prognostic markers of pancreatic ductal adenocarcinoma. Front Genet 2019 Jul 18;10:624. [doi: 10.3389/fgene.2019.00624] [Medline: 31379917]
19. Liu K, Jin M, Xiao L, Liu H, Wei S. Distinct prognostic values of mRNA expression of glutathione peroxidases in non-small cell lung cancer. Cancer Manag Res 2018 Aug;10:2997-3005. [doi: 10.2147/cmar.s163432]
20. Giannoni A, Baruah R, Leong T, Rehman MB, Pastormerlo LE, Harrell FE, et al. Do optimal prognostic thresholds in continuous physiological variables really exist? Analysis of origin of apparent thresholds, with systematic review for peak oxygen consumption, ejection fraction and BNP. PLoS One 2014;9(1):e81699 [FREE Full text] [doi: 10.1371/journal.pone.0081699] [Medline: 24475020]
21. Hess K. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Stat Med 1995 Aug 15;14(15):1707-1723. [doi: 10.1002/sim.4780141510] [Medline: 7481205]
22. Xu R, Shaw PA, Mehrotra DV. Hazard ratio estimation in small samples. Stat Biopharm Res 2018 Jun 20;10(2):139-149. [doi: 10.1080/19466315.2017.1369899]
23. Pirker R, Pereira JR, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, FLEX Study Team. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. Lancet 2009 May 02;373(9674):1525-1531. [doi: 10.1016/S0140-6736(09)60569-9] [Medline: 19410716]
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012 Oct 04;490(7418):61-70 [FREE Full text] [doi: 10.1038/nature11412] [Medline: 23000897]
25. Cochran WG. Some methods for strengthening the common χ 2 tests. Biometrics 1954 Dec;10(4):417. [doi: 10.2307/3001616]
26. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959 Apr;22(4):719-748. [doi: 10.1093/jnci/22.4.719] [Medline: 13655060]
27. Gehan EA. A generalized two-sample Wilcoxon test for doubly censored data. Biometrika 1965 Dec 01;52(3-4):650-653. [doi: 10.1093/biomet/52.3-4.650]

XSL•FO
RenderX

28. Breslow N. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 1970;57(3):579-594. [doi: 10.1093/biomet/57.3.579]

Abbreviations

FDR: false discovery rate
HR: hazard rate

Edited by R Kukafka; submitted 01.02.21; peer-reviewed by T Karn, YH Kim; comments to author 15.02.21; revised version received 19.02.21; accepted 06.05.21; published 26.07.21.

Please cite as:
Lánczky A, Győrffy B
Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation
J Med Internet Res 2021;23(7):e27633
URL: https://www.jmir.org/2021/7/e27633
doi:10.2196/27633
PMID:34309564

©András Lánczky, Balázs Győrffy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Changes in Language Style and Topics in an Online Eating Disorder Community at the Beginning of the COVID-19 Pandemic: Observational Study

Johannes Feldhege[1], MSc; Markus Moessner[1], PhD; Markus Wolf[2], PhD; Stephanie Bauer[1], PhD

[1]Center for Psychotherapy Research, Heidelberg University Hospital, Heidelberg, Germany

[2]Department of Psychology, University of Zurich, Zurich, Switzerland

Corresponding Author:
Johannes Feldhege, MSc
Center for Psychotherapy Research
Heidelberg University Hospital
Bergheimer Strasse 54
Heidelberg, 69115
Germany
Phone: 49 622156 ext 7876
Fax: 49 6221567350
Email: johannes.feldhege@med.uni-heidelberg.de

Abstract

Background: COVID-19 has affected individuals with lived experience of eating disorders (EDs), with many reporting higher psychological distress, higher prevalence of ED symptoms, and compensatory behaviors. The COVID-19 pandemic and the health and safety measures taken to contain its spread also disrupted routines and reduced access to familiar coping mechanisms, social support networks, and health care services. Social media and the ED communities on social media platforms have been an important source of support for individuals with EDs in the past. So far, it is unknown how discussions in online ED communities changed as offline support networks were disrupted and people spent more time at home in the first months of the COVID-19 pandemic.

Objective: The aim of this study is to identify changes in language content and style in an online ED community during the initial onset of the COVID-19 pandemic.

Methods: We extracted posts and their comments from the ED community on the social media website Reddit and concatenated them to comment threads. To analyze these threads, we applied top-down and bottom-up language analysis methods based on topic modeling with latent Dirichlet allocation and 13 indicators from the Linguistic Inquiry and Word Count program, respectively. Threads were split into prepandemic (before March 11, 2020) and midpandemic (after March 11, 2020) groups. Standardized mean differences were calculated to estimate change between pre- and midpandemic threads.

Results: A total of 17,715 threads (n=8772, 49.5% prepandemic threads; n=8943, 50.5% midpandemic threads) were extracted from the ED community and analyzed. The final topic model contained 21 topics. CIs excluding zero were found for standardized mean differences of 15 topics and 9 Linguistic Inquiry and Word Count categories covering themes such as ED symptoms, mental health, treatment for EDs, cognitive processing, social life, and emotions.

Conclusions: Although we observed a reduction in discussions about ED symptoms, an increase in mental health and treatment-related topics was observed at the same time. This points to a change in the focus of the ED community from promoting potentially harmful weight loss methods to bringing attention to mental health and treatments for EDs. These results together with heightened cognitive processing, increased social references, and reduced inhibition of negative emotions detected in discussions indicate a shift in the ED community toward a pro-recovery orientation.

(J Med Internet Res 2021;23(7):e28346) doi:10.2196/28346

KEYWORDS

COVID-19; eating disorders; online eating disorder community; language; mental health; social media; LIWC; Linguistic Inquiry and Word Count; Reddit; topic modeling

XSL•FO
RenderX

Introduction

COVID-19, caused by SARS-CoV-2, emerged in late 2019 in Wuhan, China and has since spread worldwide before being declared a global pandemic on March 11, 2020, by the World Health Organization. With the number of infections and deaths in the millions, the COVID-19 pandemic has already led to suffering for much of the world population. The pandemic has also contributed to elevated levels of anxiety, depression, and stress in the general population [1,2], with potentially greater effects on individuals with pre-existing mental disorders [3]. Quarantines, social distancing, lockdowns, and other public health measures taken to contain the spread of COVID-19 also have the potential for adverse effects on psychological well-being [4]. These measures were associated with increased depression, anxiety, and psychological distress in the general population [5], with indications for persisting effects after lockdowns were lifted [6]. During previous epidemics, experiences of quarantine led to long-term increases in depressive symptoms [7] and heightened anxiety symptoms, especially in those with a history of mental disorders [8]. Similarly, the COVID-19 pandemic and its associated public health measures impact individuals with lived experience of eating disorders (EDs) in numerous ways, affecting their symptomatology, social support, coping mechanisms, treatment, and engagement with media and the internet.

Individuals with EDs showed higher psychological distress including heightened fear and health anxiety since the beginning of the pandemic [9,10]. ED symptoms such as bingeing and food restriction have worsened [10-13], which has been linked to increased food insecurity, as opportunities to shop for food have been reduced and food shortages appeared in supermarkets [14]. Former patients with bulimia nervosa reported higher shape, weight, and eating concerns; body dissatisfaction; and drive for thinness [13]. Studies have also found a higher prevalence of compensatory behaviors such as purging, excessive exercise, or the abuse of laxative and diuretics during this time [9,10,12,13].

Health and safety measures due to COVID-19, such as quarantine, social distancing, or lockdowns disrupted the structure and routine of everyday life, as most people spent a lot more time at home than usual [15]. For many individuals with EDs, this has led to the loss of familiar coping mechanisms and increased rumination about food, weight, or exercise [11]. Similarly, social support has been impacted by health and safety measures that left many individuals isolated and cut off from their usual social support networks [10]. Additionally, access to health care services and ED treatment has been limited during this time [13,15,16]. Although many services transitioned to telehealth applications rather quickly, patients report that the quality of their treatment has declined during the pandemic [10]. At the same time, contacts to helplines and instant chats have increased compared to previous years with individuals contacting these services being more strongly affected by EDs, depression, or anxiety [17].

The COVID-19 pandemic also put topics such as food, weight gain, and physical exercise, which can be triggering for individuals with EDs, into the spotlight of traditional and social media [9,11]. This includes news coverage of food shortages and news reports about potential weight gain due to lower activity levels and increased time spent at home as well as a spread of online workout videos on social media [14,18].

For individuals with mental health issues, and especially for individuals with a lived experience of an ED, or those who are at risk for developing an ED, social media has been an important source of communication and social support from like-minded peers due to the shame and stigma associated with these conditions [19]. The expectation would be that social media became even more important for these individuals during the COVID-19 pandemic, a time when access to other sources of support is reduced and people have to spend more time at home. A mixed-methods study on the impact of the COVID-19 pandemic on individuals with lived experience of an ED in the United Kingdom reported that most participants are spending more time on the internet and on social media and that many felt that this had a negative effect on their ED symptoms [11]. One reason for the negative impact of time spent online might be that some participants visited potentially harmful online pro-ED communities, also called pro-ana (short for pro–anorexia nervosa) or pro-mia (short for pro–bulimia nervosa). These websites or social media channels are considered harmful because they often glorify EDs and encourage visitors to engage in pathological ED-related behaviors rather than supporting them to seek help or recovery from EDs [20,21]. Prominent features in pro-ED communities are “thinspiration” content, text or images that propagate a thin ideal, and weight loss “tips and tricks” [22,23]. Effects of visiting pro-ED communities are increases in dieting, body dissatisfaction, negative affect, drive for thinness, and disordered eating behaviors [21,24,25].

Although these communities can be harmful, they can also be a source of social support for individuals who feel that other users with lived experience of EDs can understand them better than real-life friends and family [26,27]. As a consequence, people with EDs, or those at-risk for developing an ED, could be reaching out to these communities during the COVID-19 pandemic to receive the social support they are lacking in their everyday life. A qualitative study on three ED communities on the social media website Reddit during the first months of the COVID-19 pandemic uncovered themes such as increased ED symptomatology, changes in daily routine, and treatment interruptions [28], which echo issues that were also found in survey studies.

Research on social media can complement, extend, and even offer some advantages over more traditional clinical research approaches. For example, on social media, a wide group of individuals is active including those at risk for EDs, those that have never been in treatment, or those that would not be reached by traditional surveys. Furthermore, the analysis of social media text is free from the biases introduced by experimental or interview situations. Social media research has contributed to our understanding of the social and psychological implications of disruptive events such as a global pandemic, for example, through tracking developments and interactions, and analyzing their temporal and geographical distributions [29]. Studying social media can also bring to light how an infodemic, that is,

XSL•FO
RenderX

health-related misinformation, spreads [30]. A number of phenomena related to the COVID-19 pandemic have already been explored with data from social media websites. One study charted the course of COVID-19 symptoms based on topics and language styles extracted from firsthand accounts infected individuals had shared on Reddit [31]. Another study observed the impact of lockdowns on the language styles of Twitter users in Wuhan and Lombardy [32]. The researchers discovered changes in indicators for cognitive processing, uncertainty, and increased time spent at home.

In this study, we explored how social media activity in an online ED community develops during the first months of the COVID-19 pandemic. Although this time is certainly difficult for all individuals with mental health conditions, it posed a particular challenge for individuals with lived experience of EDs. Recent survey studies have shown not only that these individuals are at high risk of experiencing deteriorating symptomatology but also that they were particularly affected by the health and safety measures imposed to curb the spread of the virus and by the news reports on food shortages, weight gain, and home workouts. We sought to determine whether the aforementioned effects such as worsened ED symptomatology and anxiety and reduction of treatment services and social support are reflected in changes in one of the largest ED communities on the social media website Reddit. Due to lockdowns and quarantines, many individuals were spending more time on the internet and social media, potentially discovering ED communities for the first time or becoming more active in them. These communities differ from other mental health online communities in that they often contain both harmful and supportive elements. Therefore, it is vital to investigate the communities that individuals with lived experience of EDs can encounter on social media and how these communities changed as the COVID-19 pandemic began to develop. To the best of our knowledge, only one qualitative study has investigated online ED communities on the social network Reddit.com at the beginning of the COVID-19 pandemic [28]. In contrast to this study, we took a quantitative approach to one of the largest ED communities on Reddit by using two state-of-the-art quantitative text analysis methods. Such methods allow researchers to turn large samples of texts into quantitative representation and to estimate differences between chosen subsets. Specifically, we applied top-down and bottom-up automated language analysis methods to track verbal behavior during the early weeks of the pandemic. First, we extracted the major topics and estimated changes in their prevalence during this time. Second, by using a validated dictionary approach, we analyzed language styles in this community before and after the initial onset of the pandemic. These two analysis methods each provide distinct insights and at the same time complement each other to produce a richer understanding of the changes in an online ED community during the first months of the COVID-19 pandemic than either method on its own. Because our study focuses on the changes in the community as a whole and not on individual users, the unit of analysis for both methods is a discussion thread, which consists of an initial post and all comments made to this post. The aim of our study was to investigate possible changes in the content and language style of these threads at the beginning of the COVID-19 pandemic.

Methods

Study Design

We conducted an observational study in an online ED community to identify changes in content and language style in comment threads after COVID-19 became a global pandemic. We chose March 11, 2020, as the start date for the global pandemic, as it coincides with the declaration of COVID-19 as a global pandemic by the World Health Organization [33]. All data were categorized as *prepandemic* (before March 11) or *midpandemic* (after March 11) in a dichotomous variable *global pandemic status*. All topic and language style variables were *z* standardized, resulting in variables with a mean of 0 and a SD of 1, allowing for easier interpretation. Changes in topics and language styles from pre- to midpandemic threads were estimated by subtracting their midpandemic mean standardized prevalence from their prepandemic mean standardized prevalence. These standardized mean differences (SMDs) between pre- and midpandemic prevalences can be considered as analogues to effect sizes such as Cohen *d*. They were illustrated in a graph together with their 99% CIs. A significant change in the mean prevalence of a topic or language style can be observed if its CI does not include zero. A stricter level of confidence at 99% was chosen as even small differences can become significant in a large data set such as the one used in this study.

Data Set

We collected data from a large ED community on the social media website Reddit. The community is not identified by name in this paper, as the users wish to remain anonymous. The community was founded in November 2017 after the largest ED community on Reddit at the time, *r/proed*, was shut down by the administrators of Reddit for violating its rules (for more information on *r/proed*, see [34]). Posts and comments in this community were accessed in regular time intervals from April 6 to May 20, 2020, through Reddit's official application programming interface (API) using the R package redditoR. As the API limits the access to 1000 items at a time, posts and comments earlier than April 6 were not available through this approach. To gather earlier posts and comments and, thus, derive a prepandemic sample, we accessed earlier posts and comments in the ED community by users who had contributed at least one post or comment in the time period between April 6 and May 20. In total, 18,071 posts and 100,143 comments created by 6683 users between November 1, 2019, and May 20, 2020, were available for analysis. We concatenated the text of a post and the texts of all comments made to that post into a single thread, thereby combining texts from different users. This approach was deemed appropriate because the focus of our study was not on changes in individual users but rather on trends in topics and language stylistics in the community as a whole. In the following, threads are used as the unit of analysis.

XSL•FO
RenderX

Data Preprocessing

We removed 195 posts and 7 comments made by self-identified bots from the data. The native language of posts and comments was determined using the R packages cld2 and cld3, and 22 comments in a language other than English were excluded from further analyses. The threads were prepared for the text analyses by removing HTML code and Unicode characters.

Topic Modeling

We used topic modeling with latent Dirichlet allocation (LDA) [35] to discover latent topics in threads. LDA is a state-of-the-art unsupervised bottom-up text analysis method that has previously provided intelligible topics for text corpora from Reddit communities for depression [36] and EDs [34]. It can be applied on large text corpora without manual coding and little input by the researchers. Two additional text preprocessing steps were performed to prepare the threads for topic modeling. First, the removal of numbers, punctuation marks, and stop words, which are common words with little meaning, such as *is* or *this*. For this, a list of stop words created by the Snowball stemmer project and included in the R package tm was used. The second preprocessing step was reducing words to their word stem. The R package stm was used to estimate a structural topic model of the preprocessed threads. The variable *global pandemic status* was included as a covariate in the model, allowing the prevalence of topics in threads to vary according to whether they were started before or after the declaration of COVID-19 as a pandemic. An initial search for the appropriate number of topics, *K*, for the corpus was conducted with topic models with different values of *K*=3, 6, 9, 12...30. In this first step, the models were evaluated using the indicators exclusivity and semantic coherence to narrow down the number of topics [37,38]. In a second step, a number of models with *K*=9 to 21 topics were estimated. For these models, we set the number of runs to 50 for each K, the number of expectation-maximization iterations to a maximum of 200, and all other parameters to default values. From these candidate models, a final model with *K*=21 topics was chosen based on inspection of semantic coherence and exclusivity and manual evaluation of interpretability of its topics. The exclusivity and semantic coherence for models in both steps are displayed in Figures S1 and S2 in Multimedia Appendix 1. Topics were manually annotated with a topic label by the main author (JF), and topic labels were reviewed by the other authors (MM, MW, and SB). Labels were chosen on the basis of the 15 most characteristic words and 20 most characteristic texts for each topic. The Results section shows the manually chosen topic labels and 7 characteristic words as measured by the FREX metric, which balances how frequent and how exclusive to one topic a word is [37].

Language Style Analysis

Language style in threads was assessed with a set of 13 indicators that have been shown to be associated with ED-related social online activities [39]. These indicators cover behavioral, affective, social, and cognitive dimensions of language use (see the Results section for the names and exemplary words for the indicators). To assess the frequency of these indicators, we used the Linguistic Inquiry and Word Count (LIWC) text analysis program [40]. LIWC is based on a word count algorithm that searches each text unit for words that are assigned to prespecified language categories in its internal dictionary. Words in a given text are matched to these categories and counted to determine the frequency of each category in the text. We also included the relative frequencies of question marks and exclamation marks provided by LIWC, as their use can be an indicator for complexity reduction in texts [39]. We excluded 161 (0.91%) threads from the analyses because less than 70% of their words were captured by the LIWC 2015 dictionary to prevent unreliable analyses (eg, short text units or due to misspellings or typing errors).

Results

The final sample used in the analysis consisted of 17,715 threads (n=8772, 49.5% prepandemic threads; n=8943, 50.5% midpandemic threads). Descriptive statistics of the threads are listed in Table 1. An average thread contained 4 (SD 7.55) comments, was populated by 4.02 (SD 5.19) users, and consisted of 258.41 (SD 390.18) words. Of the 6554 users that participated in the final sample of threads, 2595 (39.59% of all users) participated in both pre- and midpandemic threads. These users participated in more threads per day in the midpandemic period (mean 0.12, SD 0.25) than in the prepandemic period (mean 0.08, SD 0.15; t_{2594}=–9.18; P<.001). There were 3215 (49.05%) users who participated in the ED community for the first time during the midpandemic period.

Table 1. Descriptive statistics of comment threads (N=17,715) in the eating disorder community on the social media website Reddit.

Variable	Prepandemic threads	Midpandemic threads	Total
Threads, n (%)	8772 (49.5)	8943 (50.5)	17,715 (100)
Threads per day, mean (SD)	66.45 (16.70)	129.61 (43.91)	88.13 (41.74)
Users per thread[a], mean (SD)	3.43 (3.58)	4.6 (6.33)	4.02 (5.19)
Comments per thread, mean (SD)	3.24 (5.20)	4.74 (9.24)	4.00 (7.55)
LIWC[b] word count per thread, mean (SD)	204.03 (247.54)	311.76 (485.53)	258.41 (390.18)
LIWC dictionary words per thread, mean (SD)	89.83 (5.06)	90.51 (4.62)	90.17 (4.85)

[a]Users per thread includes the post author and all users that commented on the thread.

[b]LIWC: Linguistic Inquiry and Word Count.

XSL•FO
RenderX

The labels of the final topic model, exemplary words for topics and LIWC categories, and their unstandardized mean prevalence rates in threads before and after March 11, 2020, are shown in Table 2. Topics can be subsumed into broad categories such as ED symptom-related topics (*binge or restrict*, *purging*, *binge foods*, *low calorie foods*); weight, shape, and eating concerns (*weight loss or gain*, *body dysmorphia*, *exercise*, *appearance*, *meals*); mental health and treatment (*mental health*, *ED treatment*); everyday life (*domestic life*, *entertainment*, *drinks*); social aspects (*romantic relationships*, *social support*); EDs in community and society (*ED communities*, *EDs and society*); and expressions of emotions (*affect*).

Overall, the most common topics were *affect*, *support*, *time*, and *binge or restrict*. The order of topics from most to least prevalent changes from pre- to midpandemic threads, with the topics *meals* and *romantic relationships* becoming more common than the topic *binge or restrict* as an example. SMDs of topics between pre- and midpandemic threads are shown in Figure 1. A total of 15 out of the 21 topics showed a significant change, that is, they had CIs that did not include zero. Significant increases in the first 2 months of the pandemic compared to the prepandemic time period were observed in the prevalence of the following nine topics: *affect* (SMD 0.079, 99% CI 0.040-0.117), *social support* (SMD 0.180, 99% CI 0.142-0.219), *meals* (SMD 0.057, 99% CI 0.019-0.096), *romantic relationships* (SMD 0.087, 99% CI 0.048-0.125), *mental health* (SMD 0.186, 99% CI 0.148-0.225), *ED treatment* (SMD 0.049, 99% CI 0.01-0.088), *EDs and society* (SMD 0.133, 99% CI 0.095-0.172), *ED community* (SMD 0.101, 99% CI 0.061-0.139), and *development of ED* (SMD 0.091, 99% CI 0.052-0.13). The prevalence of the following six topics decreased in midpandemic threads in relation to prepandemic threads: *binge or restrict* (SMD –0.186, 99% CI –0.225 to –0.148), *purging* (SMD –0.208, 99% CI –0.246 to –0.169), *low calorie foods* (SMD –0.048, 99% CI –0.087 to –0.009), *drinks* (SMD –0.105, 99% CI –0.144 to –0.067), *binge foods* (SMD –0.099, 99% CI –0.137 to –0.06), and *appearance* (SMD –0.067, 99% CI –0.106 to –0.029). Figures of daily mean prevalence rates of all topics between November 1, 2019, and May 20, 2020, can be found in Supplement S3 in Multimedia Appendix 1.

Out of the 13 LIWC categories, 9 showed a significant change, that is, they had a 99% CI excluding zero. Five LIWC categories, *anxiety* (SMD 0.086, 99% CI 0.047-0.125), *cognitive processes* (SMD 0.18, 99% CI 0.141-0.218), *insight* (SMD 0.113, 99% CI 0.074-0.152), *social processes* (SMD 0.101, 99% CI 0.062-0.139), and *third-person singular* (SMD 0.047, 99% CI 0.008-0.086), became more frequent in mid- compared to prepandemic threads. *Question marks* (SMD –0.146, 99% CI –0.185 to –0.107); *exclamation marks* (SMD –0.043, 99% CI –0.082 to –0.004); and words from the LIWC categories *body* (SMD –0.071, 99% CI –0.109 to –0.032), *health* (SMD –0.041, 99% CI –0.079 to –0.002), *ingestion* (SMD –0.085, 99% CI –0.124 to –0.047), and *death* (SMD –0.055, 99% CI –0.094 to –0.016) were used less frequently in threads after the onset of the COVID-19 pandemic. Figures of daily mean prevalence rates of all LIWC categories between November 1, 2019, and May 20, 2020, can be found in Supplement S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. Names, exemplary words, and unstandardized prevalences of topics (n=21) and LIWC categories (n=15) in a corpus of comment threads from the ED community on the social media website Reddit.

Name	Exemplary words[a]	Prevalence in corpus[b], mean (SD)	
		Prepandemic threads	Midpandemic threads
Topics			
Affect	feel hate brain bad just like els	11.16 (5.33)	11.58 (5.49)
Social support	hope happi deserv thank proud better strong	7.2 (5.67)	8.33 (6.69)
Time	back month ve start ago sinc now	6.81 (4.12)	6.68 (4.09)
Binge/restrict	bing fast tomorrow day restrict urg christma	6.91 (6.60)	5.76 (5.68)
Meals	hungri eat meal food dinner hunger lunch	5.98 (5.33)	6.30 (5.87)
Romantic relationships	friend tell boyfriend said ask told partner	5.64 (5.26)	6.12 (5.77)
Purging	ive im throat cant ur vomit spit	6.49 (6.45)	5.22 (5.76)
Weight loss/gain	gain scale pound lbs lose weight maintain	5.42 (7.55)	5.62 (7.62)
Domestic life	home hous room kitchen money groceri car	4.21 (5.61)	4.26 (5.99)
Mental health	disord anorexia behavior valid mental ed behaviour	3.80 (4.30)	4.66 (4.86)
ED[c] treatment	doctor hospit treatment appoint medic inpati therapist	3.81 (7.15)	4.16 (7.23)
Exercise	burn sleep gym workout faint exercis dizzi	3.93 (5.87)	3.94 (6.13)
Low calorie foods	veggi vegan salad veget soup carrot sauc	3.84 (8.43)	3.45 (7.90)
Body dysmorphia	mirror attract photo skinnier compliment skinni thinner	3.7 (5.71)	3.74 (5.79)
Drinks	drink coffe tea soda coke caffein sweeten	4.08 (9.60)	3.14 (8.16)
Binge foods	cooki chocol peanut cake cream chip pizza	3.91 (7.56)	3.20 (6.81)
EDs and society	peopl cultur judg societi shame agre opinion	3.07 (4.08)	3.66 (4.69)
Entertainment	movi song hair scene mukbang video film	2.89 (6.68)	2.68 (6.21)
Appearance	cloth wear hip boob jean waist shirt	2.95 (7.43)	2.47 (6.85)
ED communities	sub reddit post subreddit pro account delet	2.15 (4.86)	2.69 (5.98)
Development of ED	sister parent grade mom school teacher mum	2.04 (2.99)	2.32 (3.29)
LIWC[d] categories			
Positive emotion	love, nice, sweet	3.41 (2.70)	3.40 (2.54)
Negative emotion	hurt, ugly, nasty	3.64 (2.66)	3.57 (2.33)
Anxiety	worried, fearful	0.58 (0.95)	0.67 (0.94)
Sadness	crying, grief, sad	0.87 (1.27)	0.83 (1.10)
Cognitive processes	cause, know, ought	13.45 (4.70)	14.26 (4.27)
Insight	think, know	2.48 (1.92)	2.69 (1.74)
Question marks	?	1.41 (3.55)	0.99 (1.99)
Exclamation marks	!	1.06 (2.85)	0.94 (2.74)
Social words	mate, talk, they	6.30 (4.57)	6.76 (4.49)
First-person singular	I, me, mine	10.08 (4.07)	10.01 (3.78)
Third-person singular	she, her, him	0.59 (1.45)	0.66 (1.51)
Body	cheek, hands, spit	1.41 (2.10)	1.28 (1.70)
Health	clinic, flu, pill	1.67 (2.13)	1.60 (1.73)
Ingestion	dish, eat, pizza	4.18 (3.72)	3.88 (3.36)
Death	bury, coffin, kill	0.12 (0.51)	0.10 (0.40)

[a]For topics, exemplary words are the most characteristic words as measured by the FREX metric [37]. For LIWC categories, exemplary words are taken from the LIWC manual [40].

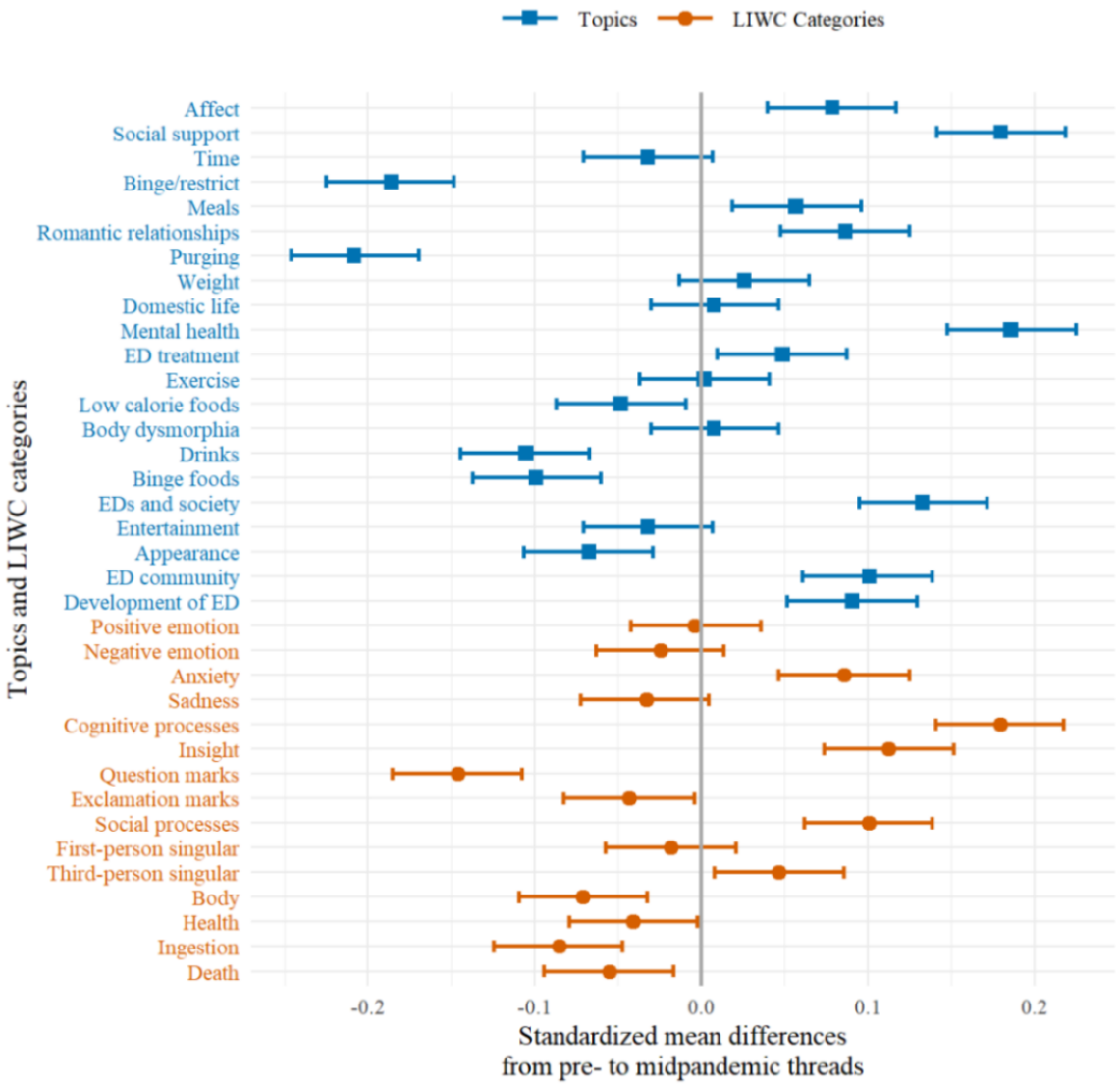
XSL•FO
RenderX

[b]For topics, prevalence in corpus is measured as the percentage of each topic in the corpus. For LIWC categories, it represents the relative percentage of category words per thread.

[c]ED: eating disorder.

[d]LIWC: Linguistic Inquiry and Word Count.

Figure 1. Standardized mean differences of the prevalence of topics and LIWC categories from pre- to midpandemic threads. ED: eating disorder; LIWC: Linguistic Inquiry and Word Count.



Discussion

Principal Findings

This study is the first to explore discussions in one of the largest ED online communities on the social media website Reddit during the onset of the global COVID-19 pandemic. We used automated text analysis methods to investigate changes in language style and topics in comment threads around the time when COVID-19 was declared a global pandemic on March 11, 2020. We were able to identify 21 topics in comment threads using LDA, addressing a number of domains and behaviors of individuals who were actively contributing to the online ED community. They cover areas such as ED symptoms; weight, shape, and eating concerns; mental health and treatment; everyday social life; or the expression of emotions. Additionally, we investigated changes in language styles at the beginning of the global pandemic using a set of indicators covering affective, cognitive, social, and behavioral dimensions of language use.

As the two text analysis methods used in our study follow different approaches, bottom-up in the case of LDA and top-down in the case of LIWC, their combined results provide a more complete picture of discussion in the ED community than either method on its own. By providing unique insights into the online social media behaviors in this large community, our results complement and contrast with existing survey studies on the effects of the pandemic on individuals with EDs (eg, [12,13]) and a qualitative analysis of other ED communities on Reddit [28]. With this study, we followed the call to study the experiences of individuals with or at risk of EDs during the global COVID-19 pandemic, as they can aid in developing studies and interventions that address the needs of these vulnerable groups in future crises [14].

XSL•FO
RenderX

Although the focus of our study was not on individual users but rather on the community as a whole, our findings were still in line with another study that showed that individuals with EDs were more active on social media during the first months of the COVID-19 pandemic [11]. Recurring users, that is, users who were active in both time periods of our study, had higher participation rates in threads in the midpandemic period. Additionally, a high number of new users joined and participated in the ED community in the midpandemic period. However, the actual number of new individuals in the ED community is most likely lower, as many Reddit users abandon their existing accounts after some time and create new accounts.

Topics covering ED symptom–related discussion such as *binge or restrict*, *purging*, *low calorie foods*, or *binge foods* had lower prevalences in midpandemic than in prepandemic threads. This was supported by a decrease in the LIWC category *ingestion* after March 11, 2020, which suggests that the communication in the community focused less frequently on ED symptom–related discussions at the onset of the global COVID-19 pandemic. This pattern stands in contrast to recent surveys that found marked increases in self-reported ED symptoms during the COVID-19 pandemic [9,11-13,16]. However, a qualitative study of ED communities during this time also noted a reduction in ED symptoms in a small percentage of users [28].

At the onset of the COVID-19 pandemic, traditional and social media were focusing on reports about food shortages, weight gain during lockdowns, and home exercises. However, this media attention did not lead to an increase in discussion about topics related to weight, exercise, and body dysmorphia in the ED community, possibly because users already show a high level of preoccupation with these topics. It is striking that communicating about physical exercise has not become more common, as surveys with individuals with an ED indicate an increase in exercising during the pandemic [11]. However, we observed a significant decrease in the prevalence of the *appearance* topic, which could indicate that users worry less about how they look at this time, most likely because they are staying at home more where they will be seen by fewer people.

Mental health and treatment-related topics became more prevalent in the ED community after the beginning of the pandemic. In their discussions, users might have used social media to address their experiences and concerns with treatment disruptions or the transition of health care services to telehealth applications [9,11,13]. However, it could also reflect a genuine effort by users to learn about ED treatments or exchange views on recovery from EDs. A willingness among users of ED communities on Reddit to foster healthier habits and attempt recovery from EDs during the first months of the pandemic was also noted in another study [28]. The increase in discussions could have also been caused by users debating or inquiring about alternatives for their interrupted ED treatments such as self-management of recovery or anonymous helplines [9,17].

On a cognitive level of language, we found a rise in indicators that were associated with cognitive processing, that is, the LIWC categories *cognitive processes* and *insight*, in threads. Cognitive processing was previously found to be elevated in blogs advocating recovery from EDs (pro-recovery) compared to pro-ED blogs [39]. This was attributed to individuals in stages of recovery showing greater cognitive reflection or reappraisal of their condition in their blogs. In line with these findings, we also observed an increased language complexity through the decreased use of exclamation marks.

Two topics demonstrating higher levels of cognitive processing were also featured more frequently in midpandemic comment threads. The topic *EDs and society* subsumes discussions on how EDs and individuals with lived experience of EDs are perceived by society as indicated by words such as *culture*, *shame*, or *judge*. With the topic *development of ED*, users reflected on events or persons in their past that they believe to have influenced the onset and development of their ED behaviors and thoughts. These discussions required a higher level of abstraction and recollection from their participants.

Users' social life seemed to play a bigger role in the community, as increases in the topic *romantic relationships*, the LIWC category *social words*, and in personal pronouns referring to another person (*third-person singular*) showed. Due to lockdowns, stay-at-home orders, and similar measures, users might be spending more time at home with their partners and families. This can be challenging for individuals with EDs, as they might feel pressured to follow a diet set by others or stressed because they are hiding their ED behaviors from others [11]. Users might be venting about social interactions or conflicts with others in their household using words from the aforementioned categories and topics. Shared meals represent another source for potential conflicts, as others can observe what and how much the individual with an ED is eating, which could have led to the increase in discussions about *meals* [11]. The amount of social references differentiated between pro-recovery and pro-ED content in a previous study, with less social references found in pro-ED texts and more social activity being associated with recovery [39]. This is attributed to a withdrawal from others or avoidance of connections with others typical for EDs. The higher frequency of social references in our study indicates that users are either having more real-world interactions about which they reflect in the online community or, alternatively, that more of their social life is happening in the community because real-life social relationships were restricted, both supporting a positive trend away from ED-typical social withdrawal.

The rise in frequency of anxiety-related words could be an indicator for elevated health anxiety and worries about contracting SARS-CoV-2 [9,10]. Indeed, health anxiety has emerged across a number of mental health communities including the ED community investigated in this study in the wake of the pandemic [41]. Additionally, anxiety about COVID-19 was related to higher ED pathology in one study [42]. In this context, it is surprising that we found a decrease in prevalence of the LIWC categories assessing *health* and *death*, as these encompass words that could be used to discuss COVID-19, its treatment, and sequelae.

Although it might appear contradictory to observe a rise in the topic *affect* without a corresponding increase in neither the *positive emotions* nor *negative emotions* categories, this

XSL•FO
RenderX

discrepancy can be explained by the composition of each category. The *affect* topic captured words representing emotions, such as shame, guilt, or anger, which are, however, featured to a much lower degree in the *positive emotions* or *negative emotions* dictionaries.

The increase in prevalence of anxiety words reflects the expression of fears associated with the serious consequences of COVID-19; however, it also points to a rise in emotional awareness and disclosure that would be more typical of recovery texts [39]. A stronger inhibition of negative emotions resulting in lower expression of negative emotion words would be expected in individuals affected by EDs rather than in individuals currently recovering from an ED. However, we did not observe any significant reductions in words from the LIWC category *sadness* in midpandemic threads.

In sum, our results illustrate pronounced changes in the online community that suggest a perspective shift in the discussions in the ED community from a narrow focus on ED symptoms and potentially harmful weight loss methods toward a broader perspective with increased attention to mental health, social resources, and treatments for EDs. This shift is characterized by heightened cognitive processing, more social references, less inhibition of negative emotions, and discussions focused less on ED symptoms and more on mental health and treatment. It is important to note here that the rules set by the community's moderators claim to not explicitly identify as either being pro-recovery or pro-ED, as the community would offer support to all and would encourage or promote neither ED behaviors nor recovery from EDs (quote paraphrased to protect the anonymity of the ED community). Although some social media platforms have distinct pro-recovery and pro-ED communities, which show seemingly little interaction between their members [43], other online ED communities feature pro-recovery and pro-ED content alongside each other [44]. It is therefore possible that the communication pattern observed in an ED community might move toward the pro-recovery end of a theoretical spectrum between pro-ED and pro-recovery. It is yet unclear whether this is a sustained change or a temporary phenomenon pushed by the disruptive situation of the recent pandemic. We also cannot deduce from these results whether they will lead to many users attempting or achieving recovery from EDs, as this can be a protracted and difficult process that is often accompanied by relapses. Remaining active in online ED communities could make achieving recovery even more difficult [44].

Other salient results from this study concern factors such as social support and the ED community. Expressions of social support became more frequent at the beginning of the pandemic. However, social support in online ED communities can be a double-edged sword, as many users report that they value the communities for the support they provide [26], while researchers suggest that support is often tied to following group norms that foster potentially harmful ED behaviors and thoughts [20]. Therefore, supportive words might be extended to users directly affected by the pandemic, to those seeking help with their condition, and potentially to those wanting to lose weight using potentially harmful methods. The topic *ED community* contains references to the ED community and Reddit in general. It could represent a kind of meta-discussion about the community and its place among other communities on Reddit with the topic. Its significant increase might reflect a kind of growth in awareness of the community and how it can support its users. These potential consequences of the pandemic on the structure of social media warrant further exploration in future research.

Limitations

This study is limited by the fact that we do not know what the relation between discussions in an ED community and real-life behaviors is. Although we found decreases in ED symptom–related discussions and increases in discussions about social and mental health issues, we do not know whether these changes are accompanied by reductions in ED behaviors and increased help-seeking or treatment uptake. However, the shift in discussion points to the community itself becoming a place that is more open to treatment and recovery from EDs.

A second limitation is that the ED community and Reddit as a whole can be used anonymously, precluding us from making inferences on the demographics or ED impairment of users in the community. Additionally, we cannot draw conclusions on changes due to the pandemic for particular users or user groups, as we were interested in changes in the community as a whole and thus estimated prevalences of language style and topics at the level of threads. As our analysis methods treat these threads, which combine a post and comments from different users as bags of words, we were unable to trace back the prevalence of topics and LIWC categories or their changes to specific users or groups of users. Although it is possible that the same users went from discussing ED symptoms to asking about mental health and treatment, it could also have been entirely new users that changed the discussion.

A third limitation concerns the time period under study. We observed the online ED community in the first few months of the pandemic, and it is yet unclear whether the observed changes will endure over time. Additionally, we cannot ascertain whether the changes we observed are fully due to the COVID-19 pandemic or whether seasonal effects also play a role. Plots of mean daily prevalence rates of topics and LIWC categories over the whole time span of our study in Supplement S3 and S4 in Multimedia Appendix 1 can give some indication whether effects are due to short spikes or more enduring trends.

Conclusions

The COVID-19 pandemic and its accompanying public health measures have disrupted the everyday life of people worldwide, possibly impacting their psychological well-being and coping strategies. Clinical researchers are understandably concerned about how these changes affect individuals with mental disorders. We present in this study a snapshot of changes in an online ED community during the first few months of the pandemic. As such, we aim to contribute to the growing area of research on the experiences of individuals affected by disordered eating during this time. Our specific contribution lies in categorizing the natural discussions occurring in the ED community on Reddit into content and language style categories and uncovering how discussions changed at the beginning of the global pandemic. The presented results suggest that reaching

XSL•FO
RenderX

out to users of online ED communities and recruiting them for treatment interventions might be especially effective at this time as need, openness, and interest for mental health treatment increases. Additionally, the language in discussions changed in a way that suggests a move toward a stronger focus on recovery and mental health treatment in the community. The changes we observed reflect issues users were experiencing in real life during the first wave of the COVID-19 pandemic. Understanding these issues can aid in developing interventions that can mitigate the consequences of future waves of the COVID-19 pandemic or other similar disease outbreaks in the future for individuals with EDs.

Acknowledgments

Preparation of this manuscript was supported by the Digital Society Initiative of the University of Zurich. The authors acknowledge financial support by the Open Access Publishing Fund of Ruprecht-Karls-Universität Heidelberg.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 1074 KB - jmir_v23i7e28346_app1.docx]

References

1. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
2. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public - A systematic review and meta-analysis. Psychiatry Res 2020 Sep;291:113190 [FREE Full text] [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
3. Neelam K, Duddu V, Anyim N, Neelam J, Lewis S. Pandemics and pre-existing mental illness: a systematic review and meta-analysis. Brain Behav Immun Health 2021 Jan;10:100177 [FREE Full text] [doi: 10.1016/j.bbih.2020.100177] [Medline: 33251527]
4. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
5. Benke C, Autenrieth LK, Asselmann E, Pané-Farré CA. Lockdown, quarantine measures, and social distancing: associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. Psychiatry Res 2020 Nov;293:113462 [FREE Full text] [doi: 10.1016/j.psychres.2020.113462] [Medline: 32987222]
6. Gan Y, Ma J, Wu J, Chen Y, Zhu H, Hall BJ. Immediate and delayed psychological effects of province-wide lockdown and personal quarantine during the COVID-19 outbreak in China. Psychol Med 2020 Aug 13:1-12 [FREE Full text] [doi: 10.1017/S0033291720003116] [Medline: 32787981]
7. Liu X, Kakade M, Fuller CJ, Fan B, Fang Y, Kong J, et al. Depression after exposure to stressful events: lessons learned from the severe acute respiratory syndrome epidemic. Compr Psychiatry 2012 Jan;53(1):15-23 [FREE Full text] [doi: 10.1016/j.comppsych.2011.02.003] [Medline: 21489421]
8. Jeong H, Yim HW, Song Y, Ki M, Min J, Cho J, et al. Mental health status of people isolated due to Middle East respiratory syndrome. Epidemiol Health 2016;38:e2016048. [doi: 10.4178/epih.e2016048] [Medline: 28196409]
9. Clark Bryan D, Macdonald P, Ambwani S, Cardi V, Rowlands K, Willmott D, et al. Exploring the ways in which COVID-19 and lockdown has affected the lives of adult patients with anorexia nervosa and their carers. Eur Eat Disord Rev 2020 Nov;28(6):826-835 [FREE Full text] [doi: 10.1002/erv.2762] [Medline: 32643844]
10. Termorshuizen JD, Watson HJ, Thornton LM, Borg S, Flatt RE, MacDermod CM, et al. Early impact of COVID-19 on individuals with self-reported eating disorders: a survey of ~1,000 individuals in the United States and the Netherlands. Int J Eat Disord 2020 Nov;53(11):1780-1790. [doi: 10.1002/eat.23353] [Medline: 32720399]
11. Branley-Bell D, Talbot C. Exploring the impact of the COVID-19 pandemic and UK lockdown on individuals with experience of eating disorders. J Eat Disord 2020;8:44 [FREE Full text] [doi: 10.1186/s40337-020-00319-y] [Medline: 32874585]
12. Phillipou A, Meyer D, Neill E, Tan EJ, Toh WL, Van Rheenen TE, et al. Eating and exercise behaviors in eating disorders and the general population during the COVID-19 pandemic in Australia: initial results from the COLLATE project. Int J Eat Disord 2020 Jul;53(7):1158-1165 [FREE Full text] [doi: 10.1002/eat.23317] [Medline: 32476163]
13. Schlegl S, Meule A, Favreau M, Voderholzer U. Bulimia nervosa in times of the COVID-19 pandemic-results from an online survey of former inpatients. Eur Eat Disord Rev 2020 Nov;28(6):847-854 [FREE Full text] [doi: 10.1002/erv.2773] [Medline: 32767858]

XSL•FO
RenderX

14. Weissman RS, Bauer S, Thomas JJ. Access to evidence-based care for eating disorders during the COVID-19 crisis. Int J Eat Disord 2020 May;53(5):369-376 [FREE Full text] [doi: 10.1002/eat.23279] [Medline: 32338400]
15. Machado PPP, Pinto-Bastos A, Ramos R, Rodrigues TF, Louro E, Gonçalves S, et al. Impact of COVID-19 lockdown measures on a cohort of eating disorders patients. J Eat Disord 2020 Nov 02;8(1):57 [FREE Full text] [doi: 10.1186/s40337-020-00340-1] [Medline: 33292539]
16. Schlegl S, Maier J, Meule A, Voderholzer U. Eating disorders in times of the COVID-19 pandemic-results from an online survey of patients with anorexia nervosa. Int J Eat Disord 2020 Nov;53(11):1791-1800 [FREE Full text] [doi: 10.1002/eat.23374] [Medline: 32841413]
17. Richardson C, Patton M, Phillips S, Paslakis G. The impact of the COVID-19 pandemic on help-seeking behaviors in individuals suffering from eating disorders and their caregivers. Gen Hosp Psychiatry 2020;67:136-140. [doi: 10.1016/j.genhosppsych.2020.10.006] [Medline: 33129138]
18. Pearl RL. Weight stigma and the "Quarantine-15". Obesity (Silver Spring) 2020 Jul;28(7):1180-1181 [FREE Full text] [doi: 10.1002/oby.22850] [Medline: 32324954]
19. Rains SA, Peterson EB, Wright KB. Communicating social support in computer-mediated contexts: a meta-analytic review of content analyses examining support messages shared online among individuals coping with illness. Commun Monogr 2015 Mar 17;82(4):403-430. [doi: 10.1080/03637751.2015.1019530]
20. Rouleau CR, von Ranson KM. Potential risks of pro-eating disorder websites. Clin Psychol Rev 2011 Jun;31(4):525-531. [doi: 10.1016/j.cpr.2010.12.005] [Medline: 21272967]
21. Harper K, Sperry S, Thompson JK. Viewership of pro-eating disorder websites: association with body image and eating disturbances. Int J Eat Disord 2008 Jan;41(1):92-95. [doi: 10.1002/eat.20408] [Medline: 17634964]
22. Borzekowski DLG, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: a content analysis of pro-eating disorder Web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
23. Juarascio AS, Shoaib A, Timko CA. Pro-eating disorder communities on social networking sites: a content analysis. Eat Disord 2010;18(5):393-407. [doi: 10.1080/10640266.2010.511918] [Medline: 20865593]
24. Peebles R, Wilson JL, Litt IF, Hardy KK, Lock JD, Mann JR, et al. Disordered eating in a digital age: eating behaviors, health, and quality of life in users of websites with pro-eating disorder content. J Med Internet Res 2012 Oct 25;14(5):e148 [FREE Full text] [doi: 10.2196/jmir.2023] [Medline: 23099628]
25. Rodgers RF, Lowy AS, Halperin DM, Franko DL. A meta-analysis examining the influence of pro-eating disorder websites on body image and eating pathology. Eur Eat Disord Rev 2016 Jan;24(1):3-8. [doi: 10.1002/erv.2390] [Medline: 26230192]
26. Csipke E, Horne O. Pro-eating disorder websites: users' opinions. Eur Eat Disord Rev 2007 May;15(3):196-206. [doi: 10.1002/erv.789] [Medline: 17676689]
27. Sowles SJ, McLeary M, Optican A, Cahn E, Krauss MJ, Fitzsimmons-Craft EE, et al. A content analysis of an online pro-eating disorder community on Reddit. Body Image 2018 Mar;24:137-144 [FREE Full text] [doi: 10.1016/j.bodyim.2018.01.001] [Medline: 29414146]
28. Nutley SK, Falise AM, Henderson R, Apostolou V, Mathews CA, Striley CW. Impact of the COVID-19 pandemic on disordered eating behavior: qualitative analysis of social media posts. JMIR Ment Health 2021 Jan 27;8(1):e26011 [FREE Full text] [doi: 10.2196/26011] [Medline: 33465035]
29. Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: a systematic review. J Biomed Inform 2016 Aug;62:1-11 [FREE Full text] [doi: 10.1016/j.jbi.2016.05.005] [Medline: 27224846]
30. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
31. Murray C, Mitchell L, Tuke J, Mackay M. Symptom extraction from the narratives of personal experiences with COVID-19 on Reddit. arXiv Preprint posted online on May 21, 2020.
32. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):4552 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
33. Listings of WHO's response to COVID-19. World Health Organization. 2020. URL: https://www.who.int/news-room/detail/29-06-2020-covidtimeline [accessed 2021-06-11]
34. Moessner M, Feldhege J, Wolf M, Bauer S. Analyzing big data in social media: text and network analyses of an eating disorder forum. Int J Eat Disord 2018 Jul;51(7):656-667. [doi: 10.1002/eat.22878] [Medline: 29746710]
35. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Machine Learning Res 2000:993-1022 [FREE Full text]
36. Feldhege J, Moessner M, Bauer S. Who says what? Content and participation characteristics in an online depression community. J Affect Disord 2020 Feb 15;263:521-527. [doi: 10.1016/j.jad.2019.11.007] [Medline: 31780138]
37. Airoldi EM, Bischof JM. Improving and evaluating topic models and other models of text. J Am Stat Assoc 2017 Jan 04;111(516):1381-1403. [doi: 10.1080/01621459.2015.1051182]

XSL•FO
RenderX

38. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 27-31, 2011; Edinburgh, Scotland, UK p. 262-272.
39. Wolf M, Theis F, Kordy H. Language use in eating disorder blogs. J Lang Soc Psychol 2013 Feb 07;32(2):212-226. [doi: 10.1177/0261927X12474278]
40. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. Austin, TX: University of Texas at Austin; 2015.
41. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
42. Scharmer C, Martinez K, Gorrell S, Reilly EE, Donahue JM, Anderson DA. Eating disorder pathology and compulsive exercise during the COVID-19 public health emergency: examining risk associated with COVID-19 anxiety and intolerance of uncertainty. Int J Eat Disord 2020 Dec;53(12):2049-2054. [doi: 10.1002/eat.23395] [Medline: 33098579]
43. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. J Med Internet Res 2012 Nov 07;14(6):e151 [FREE Full text] [doi: 10.2196/jmir.2239] [Medline: 23134671]
44. Chancellor S, Mitra T, De Choudhury M. Recovery amid pro-anorexia: analysis of recovery in social media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2111-2123 [FREE Full text] [doi: 10.1145/2858036.2858246] [Medline: 28840201]

Abbreviations

API: application programming interface
ED: eating disorder
LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
SMD: standardized mean difference

Edited by C Basch; submitted 02.03.21; peer-reviewed by C Oehler, J Chen, S Chancellor; comments to author 13.04.21; revised version received 11.05.21; accepted 18.05.21; published 08.07.21.

Please cite as:
Feldhege J, Moessner M, Wolf M, Bauer S
Changes in Language Style and Topics in an Online Eating Disorder Community at the Beginning of the COVID-19 Pandemic: Observational Study
J Med Internet Res 2021;23(7):e28346
URL: https://www.jmir.org/2021/7/e28346
doi:10.2196/28346
PMID:34101612

©Johannes Feldhege, Markus Moessner, Markus Wolf, Stephanie Bauer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 08.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Newly Developed Online Peer Support Community for Depression (Depression Connect): Qualitative Study

Dorien Smit[1,2], MSc; Janna N Vrijsen[1,3], PhD; Bart Groeneweg[4], PhD; Amber Vellinga-Dings[5], MSc; Janneke Peelen[6], PhD; Jan Spijker[1,2], MD, PhD

[1]Depression Expertise Center, Pro Persona Mental Health Care, Nijmegen, Netherlands

[2]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

[3]Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, Netherlands

[4]Dutch Depression Association, Amersfoort, Netherlands

[5]Interuniversity Center for Social Science Theory and Methodology, Department of Sociology, University of Groningen, Groningen, Netherlands

[6]Research Group for Key Factors in Youth Care, HAN University of Applied Sciences, Nijmegen, Netherlands

Corresponding Author:
Dorien Smit, MSc
Depression Expertise Center
Pro Persona Mental Health Care
PO Box 2167
Nijmegen, 6500 WC
Netherlands
Phone: 31 647074551
Email: d.smit@propersona.nl

Abstract

Background: Internet support groups enable users to provide peer support by exchanging knowledge about and experiences in coping with their illness. Several studies exploring the benefits of internet support groups for depression have found positive effects on recovery-oriented values, including empowerment. However, to date, little attention has been paid to user narratives.

Objective: This study aims to capture the user perspective on an online peer support community for depression with a focus on the modes of user engagement and the benefits users derive from participation in the forum.

Methods: In this qualitative study, we conducted 15 semistructured interviews with users of Depression Connect, a newly developed online peer support community for individuals with depression. Combining a concept-driven and a data-driven approach, we aimed to gain insight into what users value in our Depression Connect platform and whether and how the platform promotes empowerment. We performed a thematic analysis to explore the merits and demerits reported by users by using theoretical concepts widely used in internet support group research. In the subsequent data-driven analysis, we sought to understand the relationship between different styles of user engagement and the participants' experiences with the use of Depression Connect. Data analysis consisted of open, axial, and selective coding. To include as diverse perspectives as possible, we opted for purposive sampling. To verify and validate the (interim) results, we included negative cases and performed member checks.

Results: We found participation in Depression Connect contributes to a sense of belonging, emotional growth, self-efficacy, and empowerment. "Getting too caught up" was the most frequently reported negative aspect of using Depression Connect. The deployment and development of three participation styles (ie, reading, posting, and responding) affected the perceived benefits of Depression Connect use differentially, where the latter style was central to enhancing empowerment. "Being of value to others" boosted the users' belief in their personal strength. Finally, Depression Connect was predominantly used to supplement offline support and care for depression, and it mainly served as a safe environment where members could freely reflect on their coping mechanisms for depression and exchange and practice coping strategies.

Conclusions: Our findings shed new light on user engagement processes on which internet support groups rely. The online community primarily served as a virtual meeting place to practice (social) skills for deployment in the offline world. It also allowed the members to learn from each other's knowledge and experiences and explore newly gained insights and coping skills.

(J Med Internet Res 2021;23(7):e25917) doi:10.2196/25917

XSL•FO
RenderX

KEYWORDS

depression; online peer support community; internet support group; experiential knowledge; self-management; empowerment; qualitative research; patients' perspectives; participation style

Introduction

The increased accessibility of the internet, together with the advantages of offline peer support [1-3], has boosted the development of internet support groups (ISGs). These ISGs enable users to provide peer support by exchanging knowledge about and experiences of coping with a physical or mental illness [4]. Given its recurrent, persistent nature [5] and the stigma associated with depression [6], people living with the disorder often search for self-help resources [7] and appear to be the most-active users of ISGs, logging in or posting the most frequently [4,8,9].

There has been much focus on the efficacy of ISGs for depression, with previous research examining clinical outcomes and providing compelling but inconclusive evidence for a reduction of depressive symptoms resulting from the engagement in mental health ISGs (MHISGs) [4,10] or depression-specific ISGs [11]. Additionally, descriptive content analysis studies [11,12-17], user survey studies [18-21], and randomized trials or randomized controlled trials (RCTs) [22-29] evaluating ISGs for depression generally present positive results on recovery-oriented values, such as personal strength and needs and experiences with (the road to) recovery [30]. For example, content analysis studies [11,12,14-17] and user survey studies [18-21] collectively indicate that engagement in a depression ISG increases a sense of social and emotional support, and RCTs and other clinical trials suggest short-term improvements in empowerment [22], reappraisal [24], and self-efficacy [26]. However, in this body of research, the user perspective has received far less attention [18]. Such a narrative perspective on associations between processes of user engagement and the perceived value of ISG use can increase our understanding of what users need to benefit from web-based depression platforms.

Recent RCTs on MHISGs indicate that high user engagement quantified in terms of the number of posts [29] or login frequencies [31] is relevant for attaining health gains [29,31]. This, however, implies that content analysis studies may be biased. Based on the 1% rule, which postulates that 1% of users contribute around 75% of all ISG posts [32,33], content analysis studies inevitably evaluate data of small groups of highly engaged users (often referred to as "superusers" or "posters") without considering "lurkers" (users who follow discussions but seldom participate in them by posting) [34], whom we prefer to refer to as "readers." Moreover, operationalized in quantitative terms, high user engagement does not capture its qualitative nature [35]. Research into ISG participation styles does allow such a qualitative assessment, with studies revealing very diverse styles across online health communities, including ISGs [36]. As to participation styles in MHISGs, the most highly engaged users were typified as "emotionally supportive companions" [35] and "active help providers" [37,38], whereas the less-active users tended to engage more in topics regarding experiential knowledge, disclosure, and informational support [35]. Considering depression-specific ISGs, the profiles identified included "concerned about daily living," "information seekers" [19], and "interactive peer support" [20]. Moreover, contrary to quantitative analyses, qualitative characterizations of user engagement (eg, in terms of participation styles) have not yet explored how these relate to the users' valuation of the benefits and drawbacks of the platforms.

Particularly enhanced empowerment appears to play a key role [16,22,39-41] in (depression) ISGs, where gains are assumed to be linked to frequent user engagement [42-45] and, possibly, particular participation styles. However, the conceptualization of empowerment lacks clarity [42,46,47], whereas the measures to chart users' perspectives were also very diverse, both in nature and quality [42]. Based on their analysis of 17 definitions used in the literature, Cerezo et al [46] proposed the following narrow definition of *empowerment* in the context of patients with chronic illnesses such as depression: "an enabling process whereby health care professionals collaborate with patients to help them acquire knowledge and resources and whose outcome is a patient with greater ability to exercise control, manage his/her condition and to make informed decisions," precluding peer-to-peer empowerment. Empowerment is a multifaceted concept [47] and is considered both a process and an outcome, with an intrapersonal component ("sense of control"), an interactional element ("critical awareness of the sociopolitical environment"), as well as a behavioral aspect ("community involvement") [48,49]. Most studies evaluating effects of ISGs on empowerment focus on the intrapersonal component [50], whereas social processes in online communities are also likely to foster interactional empowerment [45]. Taken together, ISG use appears to promote different aspects of empowerment, but it remains unknown whether this is dependent on the nature of user engagement in relation to differential participation styles.

This study is part of a larger research project called "The Power of Depression" in which we seek to build on the recovery approach in mental health [30]. In a first exploratory study, we interviewed patients with recurrent and chronic symptoms of depression to gauge their experiential knowledge about coping strategies. The results suggested that gains in experiential knowledge mainly pertained to three intrapersonal factors: introspection, empowerment, and self-management strategies [51]. Subsequently, to facilitate the exchange of personal experiences, we developed "Depression Connect"— a closed, moderated platform providing online peer support for individuals living with depression. This platform, with a forum as its main feature, was created with the aid of a design thinking methodology following the Human Centered Design Kit (Radboudumc, REshape Center), in close collaboration with potential users currently dealing with depression, their significant others, and health professionals (psychiatrists, therapists, and psychology researchers). We made Depression Connect accessible for any person seeking help and support for depression, independent of their clinical and demographic characteristics.

XSL•FO
RenderX

We are in the process of evaluating the self-reported effects of Depression Connect on various aspects of empowerment in a quantitative longitudinal user survey (in preparation). In the qualitative evaluation we present here, we specifically sought to delineate the perceived benefits of Depression Connect participation by evaluating user experiences as a function of their participation styles. Considering the promotion of empowerment key to ISGs [16,22,39-41], as well as social and emotional dimensions that foster empowerment, we expected that participation in Depression Connect would affect users' sense of empowerment differentially depending on their mode of engagement.

Methods

Depression Connect

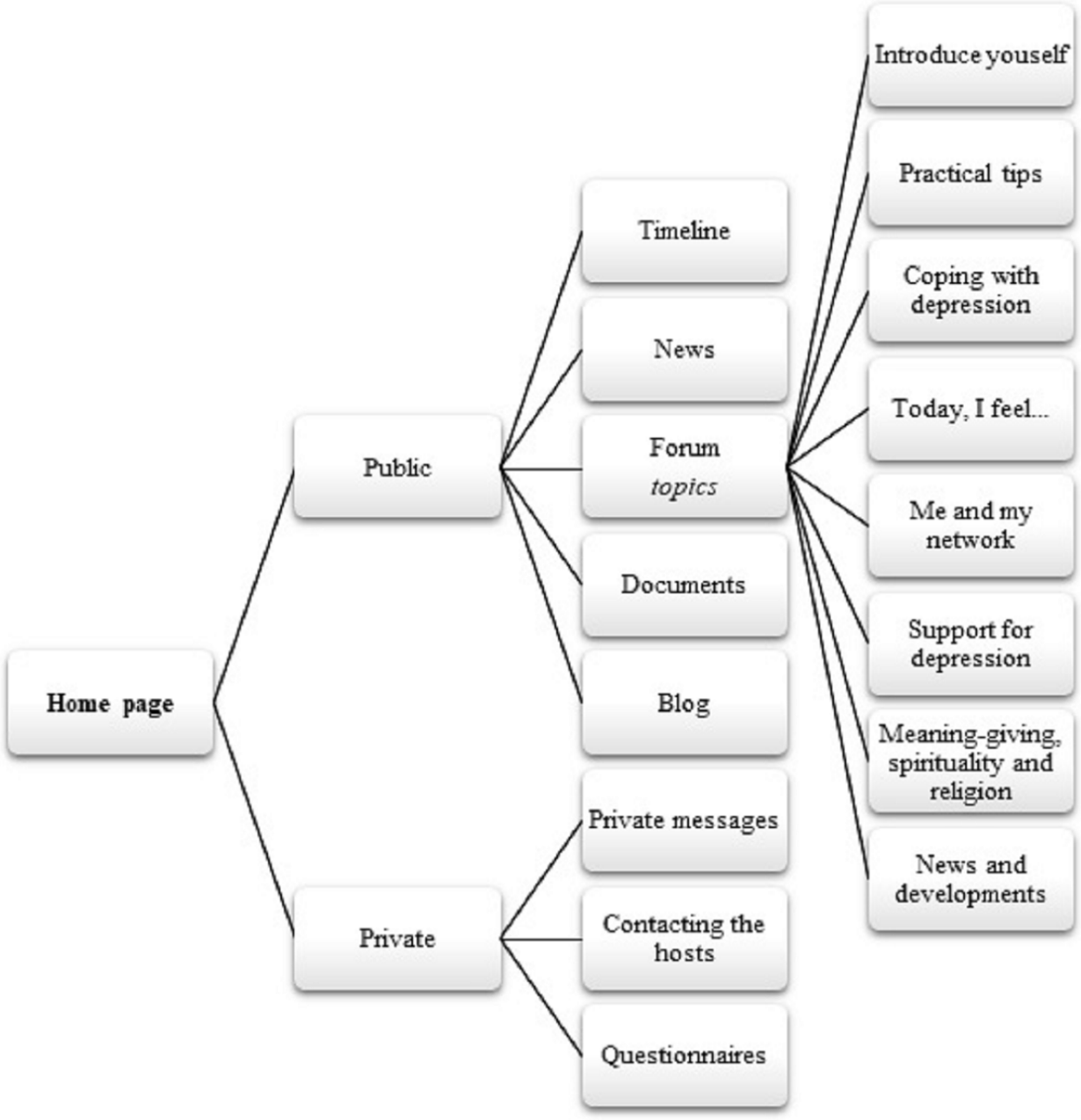
The online peer support community Depression Connect was launched on June 19, 2019. It is a digital platform that offers people with depression the opportunity to (anonymously) read or exchange knowledge about and experiences with coping with depression. It can be accessed via a website hosted by the Dutch Depression Association; the national patient association plays a central role in organizing peer support facilities for this group in the Netherlands. Through their website, any person seeking help for depression has easy access to the Depression Connect community. Depression Connect was developed and is coordinated by our research group in close collaboration with the Centre of Expertise for Depression, part of the Pro Persona Mental Health Care. To recruit a clinical population for our study, we informed members of the patient association, visitors to the website, and patients receiving treatment in a Pro Persona Mental Health Care clinic about Depression Connect and our research project through presentations, email, and flyers. We also posted the launch of Depression Connect as a news item on various websites associated with mental health care. Although other ISGs for depression are available in the Netherlands, the close collaboration between specialized mental health services and the patient association is one of the main strengths of the Depression Connect platform. When moderating and coordinating the Depression Connect community, the perspectives of both health professionals and experiential experts are taken into account. Moreover, its structural embedding in the patient and professional organizations fosters topical relevance. For example, by posting news items about depression, both organizations can inspire conversations among users and serve as a reference framework inducing users to revisit the platform regularly.

Next, we outline the login procedure, guidelines for moderators, and functionalities of Depression Connect. When accessing the site, general terms and conditions for users, privacy policy, and engagement rules are displayed; this information can also be accessed from the homepage at all times. Any interested user can then sign up for Depression Connect membership. To access the content of the community, members always need to login. When they do so for the first time, they are invited to introduce themselves; this is not mandatory and anonymized profiles are allowed. However, the moderators can always access personal contact details (name and email) to reach members personally, if necessary. Subsequently, new users will see a manual explaining how to use Depression Connect. Upon posting the first message, users are welcomed by a member of the Depression Connect moderator team. In order to ensure a constructive exchange of peer-to-peer experiences, posts are screened twice a day by one of the 5 moderators. Since the focus groups informing the development of Depression Connect expressed a clear need for *peer* support without the involvement of professionals, experiential experts were recruited as moderators. At the end of their (morning or evening) shift, the moderators document peculiarities and the general atmosphere in the forum in a logbook to inform their successor. Moderators only intervene when the content discussed, or a member's conduct, gives rise to conflicts with engagement rules. More specifically, they will act only when an urgent request for support is posted or when they identify suicidal tendencies in posts, and when rules of engagement are violated (eg, when contributors show disrespect for one another, disclose sensitive information to Depression Connect nonmembers or outsiders, share information on suicide, or share privacy-sensitive information such as names of doctors). When users exchange misinformation about depression, moderators will refer them to reliable, evidence-based sources of information. As an extra security mechanism, predetermined trigger words, which refer to a crisis situation, will automatically generate a notification in the moderators' mailbox. Launched in mid-2019, the online community attracted an average of 88 new members a month and totaled 1374 members as of September 24, 2020, when the data for our quantitative user survey study were extracted.

The design of the overall Depression Connect website and its forum is straightforward and user friendly, promoting positive user experiences and allowing users to navigate freely [52]. Figure 1 depicts the structural organization of the Depression Connect platform. Users can create their own topics on the forum, but we also provide eight predetermined topics that we derived from the main themes of experiential knowledge identified in our first study. Besides their contribution to the forum, members can, among other options, read news items and publications about depression (posted by the Depression Connect team), post blogs, and send private messages to other Depression Connect users.

XSL•FO
RenderX

Figure 1. Content and structure of the online peer support community Depression Connect.



Study Design

In this qualitative study, semistructured interviews (see Multimedia Appendix 1 for interview guide) were conducted with Depression Connect users to explore what the online community had offered them in terms of ways to cope with their current, past, or subclinical depression. We used a hybrid approach [53] combining deductive and inductive reasoning. In the theoretical context of our larger project [51], we created a guiding framework [54] based on an inventory of experiential knowledge and the relevant literature on empowerment to deduce all relevant factors involved in the broad and complex interplay of depression ISGs [14]. We applied thematic analysis [54] to identify, examine, and gain insight into the patterns of predetermined themes. Next, using an inductive, data-driven approach based on the grounded theory [55] and Strauss' exposition of the core principles of qualitative research in social sciences [56], we kept an open mind to avoid excluding potentially relevant observations. We used specific guidelines to analyze the data; these included open, axial, and selective coding and matrices [57], as well as tree diagrams [58], drawn from the grounded theory. This comparative and iterative approach enabled us to simultaneously analyze and gather new data to further explore and integrate concepts emerging during data collection, which continued until no new main themes emerged.

By combining these top-down and bottom-up approaches, we sought not only to learn what Depression Connect users do and do not appreciate about the platform (charting both differences and similarities among users) but also to further study the role of ISGs in developing experiential knowledge in general and empowerment in particular to (in)validate existing theories. The outcomes would complement our quantitative companion study of the effectiveness of the Depression Connect community

XSL•FO
RenderX

regarding empowerment (and other aspects). In this study, all Depression Connect users were invited to complete questionnaires 3 days after enlisting, with two follow-up assessments at 3 and 6 months.

After having evaluated the research protocol in accordance with the Dutch Medical Research (Human Subjects) Act, the local ethics committee (Commissie Mensgebonden Onderzoek Arnhem-Nijmegen) waived ethical approval given the minimal burden to the study participants. All participants were asked to provide written, informed consent prior to the interview following the Declaration of Helsinki.

Study Participants

We posted three calls for participation in our study over a 3-month period in the news section of the Depression Connect platform. Eight potential participants responded. We sent them an information letter by email, inviting them for a telephone screening. During this call, the researcher provided the candidate with a brief introduction to the study and information on the purpose of the interview, explaining the voluntary nature and confidentiality of their participation. The candidates' demographic and clinical characteristics and patterns of use of the online community were assessed to ensure diversity within the sample. In order to obtain as wide a range of user perspectives as possible, we adopted lenient inclusion and exclusion criteria, resulting in all 8 candidates being included, with sufficient differences in characteristics and backgrounds.

After an initial analysis of these first eight interviews to derive the concepts discussed, we used purposive sampling to identify new participants with different profiles and uncover any additional themes. Members with (prior) experience in offline peer communities organized by the Dutch Depression Association were contacted by their regional coordinator, which yielded 1 participant. At this point, the research sample (n=9) solely consisted of individuals with recurrent or chronic depression. Therefore, a member who had newly joined Depression Connect, introducing herself as having been recently diagnosed with depression, was invited to participate via a personal email. We wanted to also include negative cases, that is, Depression Connect users with experiences or perspectives who were likely to deviate from other users and the main theories or evidence on ISG [59], to potentially provide unexpected findings that might ultimately strengthen the theory. Hence, we recruited 2 participants who distinguished themselves by their minimal or nonuse of Depression Connect after joining the platform. One of these participants enrolled himself upon our invitation, identifying himself as a Depression Connect member who mainly engaged in other online fora about depression. The second (female) participant was a former Depression Connect user randomly selected from a contact list of unsubscribed members who was invited via email. Recruited through a fourth and final call for participation on the platform, another 3 participants were interviewed to achieve data saturation, resulting in a final sample of 15 (former) Depression Connect members.

Table 1 shows the demographic and clinical characteristics of the study participants and the frequency and duration of their use of Depression Connect. All 15 participants had received some form of psychological care or treatment at an earlier stage in their lives, with 10 (67%) receiving current and 3 (20%) awaiting treatment for their depression (including 1 negative case); 2 (13%) participants (including 1 negative case) were not being treated at the time of the interview. Furthermore, 12 of 15 (80%) participants were taking or had taken psychotropic agents for their depression. The majority (11/15, 73%) visited the Depression Connect forum regularly, with a frequency varying from daily to once a week, barring, by definition, the 2 negative cases and 2 other members who joined the forum only irregularly.

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of study participants (N=15) and their engagement on the Depression Connect online support community.

Demographic characteristics	Value
Age (years), mean (SD)	49 (11)
Gender, n (%)	
Male	6 (40)
Female	9 (60)
Ethnicity, n (%)	
Caucasian of Dutch descent	15 (100)
Educational level, n (%)	
Secondary education (middle or high school)	2 (13)
Secondary vocational education and training	7 (47)
Advanced vocational education and training and academic education	6 (40)
Clinical characteristics, n (%)	
Current mental health care or treatment	
Intake or waiting list	3 (20)
Ongoing	10 (67)
Mental-health nurse practitioner (general practice)	3 (20)
Psychologist or psychotherapist (secondary care)	7 (47)
None	2 (13)
Treatment history[a]	
Secondary mental health care (eg, CBT[b], psychotherapy)	15 (100)
Previous psychopharmacological treatment	6 (40)
Current psychopharmacological treatment	8 (53)
Never used psychotropic medication	1 (7)
Number of depressive episodes[c]	
1	1 (7)
2	1 (7)
3-5	5 (33)
Chronic course only (symptoms persisting ≥2 years)	7 (47)
Chronic course in addition to depressive episodes	3 (20)
Age at depression onset (in years; range: 8-57)	
<12	1 (7)
12-18	3 (20)
19-32	6 (40)
33-45	4 (27)
>46	1 (7)
Duration since onset (in years; range: 5-47)	
0-10	6 (40)
11-20	2 (13)
21-30	2 (13)
31-40	2 (13)
41-50	3 (20)
User engagement	

XSL•FO
RenderX

Demographic characteristics	Value
Frequency of using Depression Connect, n (%)	
Daily	6 (40)
3 times a week	4 (27)
Once a week	1 (7)
Irregular	2 (13)
Unsubscribed after 1 month of forum use (negative case)	1 (7)
Inactive (negative case)	1 (0.7)
Duration of use (in months; excluding negative cases; range: 1.5-11), mean (SD)	6.8 (3.8)

[a]Includes overlap in different forms of treatment.

[b]CBT: cognitive behavioral therapy.

[c]Includes overlap in chronic course and depressive episodes.

Data Collection

From February 2020 until June 2020, two authors (DS and AD) individually conducted semistructured interviews (n=9 and n=6, respectively) with 15 Depression Connect users (including 1 former user), lasting 28.37 to 66.16 minutes (mean 48.5, SD 11.25 minutes). Both authors have a master's degree in social sciences and are specifically trained and experienced in qualitative research methods. They had created a topic list, building upon the first exploratory study [51], the existing literature, feedback from the project group members (1 psychiatrist, 3 experiential experts on depression, and 2 senior researchers) and an exploratory interview with a Depression Connect member. As shown in Textbox 1, the following topics guided the interviews: (1) forum use (why, when, and how), (2) Depression Connect benefits and downsides, (3) Depression Connect working mechanisms, and (4) (relationship with) the use of other forms of formal or informal depression support and care. The complete interview guide is available in Multimedia Appendix 1. Based on interim analyses conducted after four and eight interviews, DS and AD reviewed the topic list and incorporated newly identified themes. First, we formulated new questions inquiring into the perceived associations between forum use and personal recovery (coping with depression in daily life), social recovery (effect of social ties and activities), and clinical recovery. Second, to further delineate the effects of Depression Connect use, we added questions about the development and deployment of participation styles. The adjusted topic list was then used for data collection in the successive interviews [60].

Data collection took place during the COVID-19 pandemic. The Netherlands was in the early stages of the COVID-19 outbreak when we conducted the first three face-to-face interviews. Consistent with the national measures at the time, the interviewer and participants washed their hands and maintained a physical distance of 5 feet. When new COVID-19 measures stipulated that social contact be limited, the subsequent 12 interviews were conducted via video calls. All interviews were audio-recorded and transcribed verbatim, omitting any potentially identifying data.

Textbox 1. Depression Connect interview themes and subthemes.

- Use of the online community
  - Reason(s) for subscribing
  - When, why, and how is Depression Connect used
- Merits and demerits
  - Effect of Depression Connect on coping and living with depression: practical skills, meaning-giving, personal development (self-reflection)
- Working mechanisms
  - Ways in which Depression Connect as an online community and peer support method exerts its effects
- Context: other support or care
  - Ways in which Depression Connect as an online community and peer support method exerts its effects

Analysis Strategy

The data were analyzed in ATLAS.ti (version 8.4; Scientific Software Development GmbH). Given our deductive–inductive approach, coding was both concept-driven and open. For the deductive analysis, we prepared a priori thematic codes capturing relevant themes based on the research aim and topic list. To allow findings to emerge from frequent themes without restraints imposed by predetermined concepts [58], we used open, axial, and selective coding in the inductive analysis

XSL•FO
RenderX

[56,57]. To avoid a very narrow perspective, each interview started on an open-coded basis. The data were disassembled into fragments, which were compared with each other and grouped into subject categories. We used a hierarchical category system (eg, a tree diagram [57] to indicate subordinate and parallel codes and categories. When no new open codes were necessary to cover the data, axial coding was initiated. This more abstract process was used to find connections between and among categories and give coherence to the emerging analysis. Dominant and less-important elements in the data were determined to allow selective coding. At that point, the inductive and deductive approaches were combined by harmonizing the category system (based on open and axial coding) with the predetermined concepts (eg, empowerment) [51]. Categories were thus organized and integrated to uncover relationships between user engagement, Depression Connect appreciation, and the working mechanisms Depression Connect members had proposed. An open network, not specifically indicating causal linkages [58], was developed in which all the data, including the negative cases, was described and interpreted.

To ensure interrater reliability, authors DS and AD met at each stage of the process to discuss codes and themes and resolve any discrepancies. Coding was performed by an independent researcher experienced in qualitative research but not involved in the research project. The small intercoder variance was resolved by analyzing the coded segments collectively. To increase analytic sensitivity, inconsistently coded blocks were segmented into smaller units and awarded a more specific code, accompanied with a definition that included criteria for the coding of similar segments [61]. Potential interviewer or researcher bias was reduced by having participants check the outcomes to validate and verify the interim and end results. At the first member check after seven interviews, we sent all 7 participants a synthesized summary of the data analyzed thus far by email to verify whether the results resonated with their individual experiences. Participants were asked to read, comment, and return the forms. We used nonscientific wording and open questions, leaving room for individual feedback. Six participants returned the forms, and their responses were incorporated into the data set to match this data to the open network [62]. At the second member check after the final interview, we sent all 15 participants a report of the interim results together with an invitation to discuss the report per email, individual video call, or telephone. Three participants responded, providing feedback via individual video calls. Together, this enabled us to fine-tune the terminology in the interim and final results. Finally, to increase validity and to ensure any new insights into the concepts and results would be taken into account, authors DS and AD maintained a logbook in which they shared personal and theoretical views related to the research and interpretation of the data.

Participants were anonymized and identified by a randomized number (P1, P2, etc), their gender, and age. Below, we present anonymized quotes from participants to illustrate emergent themes.

Results

The interviews provided rich data covering many aspects of engagement on Depression Connect and its perceived benefits and drawbacks. We have presented the results in the order in which topics were addressed, starting with the participants' reasons to subscribe, followed by participation styles, and user valuation. We then describe the associations observed between participation styles and the perceived value of Depression Connect. Next, we summarize the negative aspects of Depression Connect use and, finally, discuss the use of Depression Connect in relation to face-to-face support, social networks, and mental health care.

Participants' Reasons to Subscribe

Given their persisting symptoms, the participants were at a stage of learning to cope and live with depression in the longer term with a focus on rehabilitation (except for 1 participant who was first diagnosed with depression 6 weeks before the interview). A total of 13 (87%) participants described a sense of loneliness or lack of social support as the main reason to engage in the online community. Their primary objective was to look for support in living with depression, which was described as a need for recognition and a genuine understanding from peers:

> *I feel quite lonely in this world. At home, it's difficult for me to speak openly about my problems. When I use the online community, I come into contact with like-minded people. Usually, for tips or a "pat on the back," things I miss at home.* [P11, male, 55 years]

One participant (negative case) emphasized this finding, while she did experience social support in daily life and unsubscribed from Depression Connect.

Participation Styles

The participants used three different participation styles: *reading* messages of peers, *posting* messages to share experiences and ask questions about (coping with) depression, and *responding* to experiences or questions of other users in order to support them. Two female participants did not post any messages because they had issues with sharing personal information. One male user did not post any responses because he struggled empathizing with fellow users.

The data show that the deployment of a specific participation style was dependent on the participants' current mood or state of mind. When feeling low, users mainly read posts or posted messages but did not respond to others' input. Overall, after joining Depression Connect, most participants first looked for support and recognition by reading the experiences from peers and posting questions about handling the illness or writing down their own story. Gradually, when their mood had improved or when they felt more at home with or committed to the Depression Connect community, participants felt more able to support their peers and started responding to others. A user's participation style could vary within a single session or differ per session, with their engagement on the forum generally developing from reading only to posting, and eventually responding:

XSL•FO
RenderX

> *At first, I thought people were just nagging a lot in their messages on the forum. I was trying to focus on solving my own problems until I saw that users were helping each other. I realized I could also benefit from their support. I began typing up my personal story. I got positive replies and then also started to respond to others.* [P14, female, 62 years]

User Valuation

Overview

In general, the participants did not report any improvements in depressive symptoms directly associated with the use of Depression Connect but often spoke of a process toward accepting the long-term nature of their depression. Hence, the values of Depression Connect lay more in the social, emotional, and practical support in learning how to manage the illness:

> *It feels good when I find recognition in the messages of others. It doesn't mean I no longer feel depressed. It just has a positive effect. Also, I get new ideas about treatment options, for example, which will eventually have a positive impact on my symptoms.* [P12, female, 47 years]

The positive effects the 15 users associated with their use of Depression Connect can be clustered into four main themes. Ranked according to their importance, these include a sense of belonging, emotional growth, self-efficacy, and empowerment.

Sense of Belonging

Most participants reported that the main benefit of Depression Connect use was the sense of belonging it provided. Recognition, emotional support, and more intrinsic understanding from peers corresponded to their reasons to subscribe, such as loneliness or a lack of support in coping with depression:

> *It feels like a warmhearted environment. You feel connected with people through recognition. Other users recognized the feelings I'm struggling with. In turn, I recognized the struggles of others in expressing and sharing their emotions. It all contributed to a natural sense of connectedness, which grew very fast. It feels like I'm in the right place.* [P2, male, 65 years]

Two participants (negative cases) did not derive a sense of belonging from the online community because they did not aim for social support; one of them felt sufficiently supported by face-to-face peer contact and the other, by her offline social network.

Emotional Growth

The data further showed Depression Connect to function as a tool for emotional growth: most participants saw Depression Connect use as an incentive to develop and reflect upon personal coping skills and ideas about the (longer-term) management of their depression. Although some topics they read about directly created a sense of recognition for a few participants (which was associated with a sense of belonging), other issues did not directly relate to them but often did trigger them into reflecting on the role the issue might or should play in the management of their depression. This process of personal identification raised the users' (self-)awareness, a necessity for the development of self-knowledge and encouragement for emotional growth. Their narratives indicated that the various processes of self-reflection encouraged them to put their problems into perspective, promoting emotion regulation:

> *Well, when you're sharing experiences you get different viewpoint and more insight, this makes you think more seriously, like, "Ah, that could be the same for me, or maybe that's a pitfall for me too." Quiet introspection can help make things more clear and may even be very helpful.* [P7, female, 53 years]

Maintaining online contact with peers did not solely serve as an incentive to reflect upon management and coping strategies. In and of itself, peer contact also helped users develop (better) communication skills. Participation on Depression Connect lowered the threshold to talk (ie, post) about depression; for some participants, the forum also served as a place to practice opening up about depression in face-to-face contacts. Moreover, disclosures tended to invite peers to challenge their negative-thinking patterns. In this context, adopting a relatively mild attitude toward oneself was mentioned as an important aspect of information sharing:

> *Maybe, it'll also become easier to speak openly to people in person. I think it's important to practice first, to really get the sense that I'm able to open up before actually doing so in more difficult situations.* [P12, female, 47 years]

Since the two negative cases did not mention emotional growth, we speculate that Depression Connect users need to experience a sense of belonging (which they also said they lacked) before they could benefit emotionally from their contact with peers.

Self-Efficacy

Most participants derived a greater sense of self-efficacy in coping with depression from the online community. Being informed or reminded about (other) coping mechanisms seemed to contribute to their sense of autonomy. Given the longer-term nature of their symptoms, users appreciated tips and experiences about specific treatments, medications, and publications on (coping with) depression the most. About half of the participants (7/15, 47%) also valued more practical advice, using the tips and recommendations about everyday activities as an incentive to (re)engage in these so-called self-management strategies, such as going for a walk or performing relaxation exercises:

> *Sometimes I read messages other users post, like "I really have to go outside more, but I don't want to," and then, a few hours later, the same user wrote "Actually, I went for a bike ride." That is when I think, "Yeah, I have to go outside too [laughs]." So yes, I have to admit, reading such posts can be an incentive. Also, certain books that people mention can make me curious, prompting me to look for more information. But it depends on how people write about things. When they share information about coping strategies, I "cherry-pick" the things that suit me most.* [P8, female, 61 years]

XSL•FO
RenderX

Empowerment

Besides increasing their sense of self-efficacy or, more specifically, autonomy, the data suggest that participating in the online community empowered most participants to come to terms with and manage their depression, with three-fourths (11/15, 73%) of our participants describing Depression Connect as a tool to provide meaning to their experiences. They explained that being of value to others living with depression and supporting peers through sharing their own experiences, provided them with a (great) sense of fulfillment:

> *What I try to convey is: Maybe you don't have any perspective now, I understand, I felt the same: "What am I doing here, on this planet?" But it will pass, really, it will pass. Even when the response is just a "thank you," it gives me fulfillment.* [P8, female, 61 years]

After all the problems they were facing because of their depression, the users felt that participating in Depression Connect *finally* afforded them a positive and valuable experience. This seems to enhance the belief in their own strength:

> *You don't get stuck in fear. For example, when you have anxieties or feel depressed, you can feel helpless, you feel lost. When you read messages of your peers saying, "It will pass," it's like, "Yes, it will." This way you encourage yourself to adopt a different attitude toward depression. And, as a consequence, when you get to feel more in balance, you can support others too.* [P10, male, 62 years]

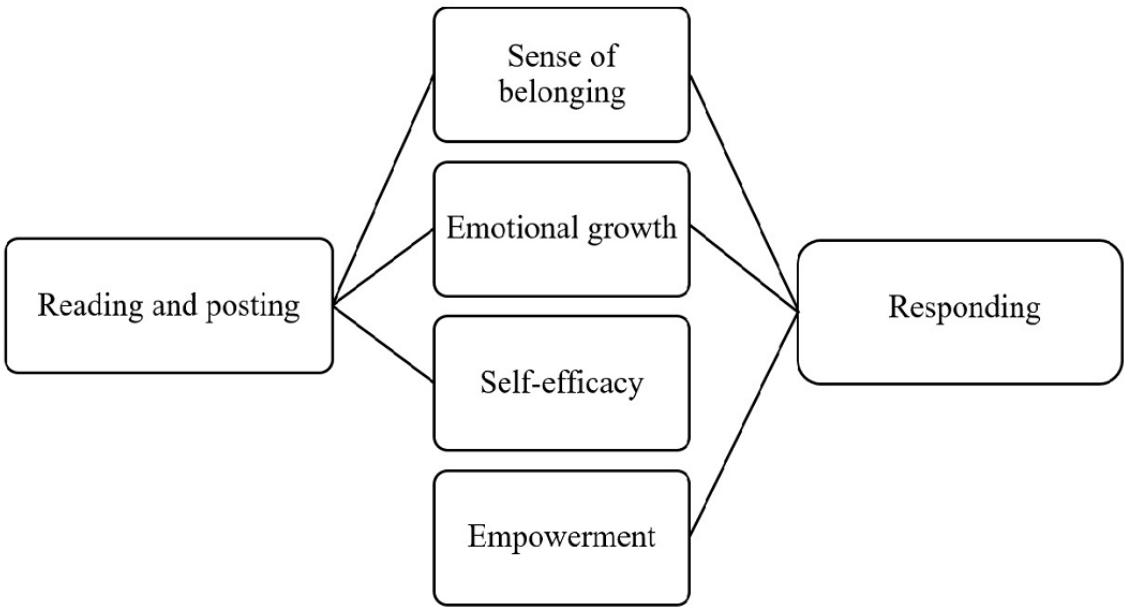
Consistent with their missing a genuine sense of connectedness with their Depression Connect peers, 2 participants (both negative cases) did not report deriving fulfillment from being of value to others.

Participation Styles and the Perceived Value of Depression Connect

Overview

To determine how and why Depression Connect users rated the merits of forum participation, we analyzed the interaction and synergy between their participation styles and valuations. The data (schematically depicted in Figure 2) suggest that reading and posting—the styles most users restricted themselves to initially—contribute to a sense of belonging, emotional growth, and self-efficacy, whereas responding, which they later engaged in, was more likely to promote empowerment in addition to a sense of belonging and emotional growth. The elucidation and participant quotes below illustrate this relationship.

Figure 2. Participation styles and the perceived value of Depression Connect.



Reading

By reading others' posts, users learned they were not alone in their struggle to learn to cope with negative feelings, recognition of which promoted a sense of belonging:

> *When others wrote about the difficulties at work they experienced on account of their depression, for instance. Suddenly, there was this recognition. A positive sensation because it made me feel like "Ah, I'm not the only one!"; it's like I read what I could have written myself…It's reassuring. Like in, company in distress makes trouble less.* [P12, female, 47 years]

Furthermore, reading their peers' experiences made users reflect on what the topics meant for them personally, furthering their emotional growth. Finally, the practical tips helped them apply (new) coping skills in their daily lives, which enhanced their sense of self-efficacy.

XSL•FO
RenderX

Posting

In general, posting played a significant and positive role in the appreciation of the forum. Writing down their feelings and struggles in managing their depression often offered users relief, whereas peer recognition and understanding or merely the knowledge that their posts were read by others was of (great) value to Depression Connect users, contributing to their development of a sense of belonging:

> *The online community serves as a "lifeline" for me. Several times, when I was really struggling, I posted a message on DC [Depression Connect]. Not to get a response, but primarily to be able to express myself by writing down my feelings. I also write for myself, to give words to my emotions. However, writing on the forum differs because I know my posts are being read. Actually, most of the time, people even respond. It's mainly the recognition they articulate that affects me, in a positive sense. Which in itself is quite strange because the recognition of others doesn't essentially change how I feel. But apparently, it works. In the sense that it sort of works as a "lifeline." A couple of times when I was doing terrible in the morning and I posted something, it was the responses of others that helped me get through the day.* [P3, male, 48 years]

Moreover, participants explained that sharing their personal story was healing. They managed to organize their thoughts when writing, often reinforcing self-reflection and emotional growth:

> *Writing about my emotions gives me peace of mind. The negative feelings don't disappear completely, but I'm better able to dissociate myself from my problems. What I get from when other users respond to my post is a sense of "not being alone in this world." More people are struggling with these same problems, who even try to help others. … It helps me put things into perspective, makes my problems feel less overwhelming.* [P4, male, 31 years]

Finally, posting specific questions about coping with depression often prompted practical tips and other new information, which is likely to have fostered a (greater) sense of self-efficacy.

Responding

Users responding to others' posts derived emotional support and recognition from their peers, which strengthened their sense of belonging. Since providing support or advice entailed having to write down one's thoughts and thus reflect on one's own experiences, this interactive participation style seemed to promote emotional growth. Similar to posting, communicating with peers helped users to better organize and formulate their thoughts:

> *Yes, yes, and I think that's exactly what the online community contributes: learning to think about, learning to reflect on yourself in other, somewhat different contexts. The things you're saying to the other are actually the things that you would like to say to yourself at that moment. Yes, and maybe that's precisely what you do, unconsciously? When you're able to sort of put yourself in the emotional world of another, then you actually feel how good it is to connect with your own emotions. This is when I realize that that is the ultimate goal.* [P2, male, 65 years]

Moreover, responding and helping peers raised the users' sense of fulfillment, which fostered a greater sense of empowerment because they felt they were of value to others.

Negative Aspects of Depression Connect Use

As to the negative aspects of Depression Connect use, these are best captured under the notion "getting too caught up." Participants explained they could become overwhelmed by the sheer volume of information appearing on the online community, the pressure of having to be continuously available and the stress caused by the concerns they had about the worries of their peers:

> *Well, you can feel overwhelmed by it all. I had mixed feelings. On the one hand, I felt relief because I could share my experiences. But on the other hand – since I visited DC [Depression Connect] a few times a day, and partly because of all the notifications I received, all the new posts – I thought, "This is not good. I'm too preoccupied with the forum and worry too much about others right now."* [P8, female, 61 years]

Users also did not appreciate the forum when their messages appeared to be misinterpreted or when they received unsolicited advice. Although, as mentioned above, half of the participants valued the practical tips on coping with depression, the other half could get frustrated because they felt they were "already aware" of the recommended strategies, or it aggravated their self-criticism because they failed to engage in the suggested activities:

> *In itself, it was good advice, definitely well-intentioned. Also, the content was completely accurate, but I was unable to follow up on it. I felt frustrated because I agreed and knew it was sane advice, it would be the sensible thing to do, but I just couldn't.* [P3, male, 48 years]

Moreover, participants reported that some members confused their own experiences and emotional needs with the personal and unique needs of peers, resulting in useless feedback and a general sense of lack of support.

Use of Depression Connect in Relation to Face-to-Face Support, Social Networks, and Mental Health Care

In general, the participants characterized the use of the online community as being complementary to their real-life peer contacts, their social network, and any professional care or treatment. As they did not feel judged by their Depression Connect peers, the participants referred to the online community as "an emotionally safe context." Not wanting to (over)burden their family and friends with their troubles, initially sharing feelings and receiving peer support online was helpful to some degree:

> *It's about the feelings you share; we're all struggling with depression. It's different from friends of mine who also suffered from depression and are the most*

XSL•FO
RenderX

> *approachable people in my network, where I sometimes think, "I don't want to bother them with my complaints again." This is much more anonymous. It is voluntary, which is nice because a friend can try to be too supportive and say, "I'll come and see you tomorrow," where I think "You don't have to come, I only felt like sharing my thoughts because I was having a bad day." Obviously, things like that don't happen in an online community like this.* [P12, female, 47 years]

At a later stage, Depression Connect interactions served as an exercise for self-disclosure in the offline world. Furthermore, anonymous participation, the voluntary nature of Depression Connect engagement, and its 24/7 availability were also mentioned as distinctive positive features compared to seeking or receiving face-to-face support via social networks or from mental health professionals:

> *The fact that you can log on day and night, that's its great strength. As opposed to my psychiatrist, whom I can't email in the middle of the night. I mean, I can, but there's no response.* [P9, female, 42 years]

The overarching principle in the relationship between the use of Depression Connect and other forms of support for depression appeared to be the opportunity the forum offered to reflect on and practice the (social) skills the users were trying to master in their daily lives or through psychotherapy. Specifically, discussing topics concerning social interactions and behavioral patterns with peers were considered beneficial:

> *I see the online community as a stepping stone for real-life social interactions with others. I learn by writing down how I should respond, how others might respond. So I'm practicing and learning. Also, I'm learning to become more self-confident so that I can connect better with others.* [P4, male, 31 years]

> *When I'm doing schema therapy with my therapist, difficult issues come to light. I found it helpful to write about these difficulties. It allows me to reflect a bit more on them, and on top of that, I can get some advice.* [P9, female, 42 years]

Discussion

Principal Findings

Central Aim

In light of the promising evidence for depression ISGs [3,11,63,64], we evaluated the relevance of Depression Connect, a newly launched online peer support community based on interviews with a selection of its users. In line with previous research [16,22], we expected that the user narratives would reflect improved empowerment. Given its central role in (online) peer support [39-41] and to clarify the concept [42,46,47], we explored the purport of its constituent constructs, and, most importantly, the role different styles of user engagement played in the users' evaluation of Depression Connect.

Perceived Value, Participation Styles, and a Central Drawback of Depression Connect Use

Participation in the online community engendered a sense of belonging and promoted the users' emotional growth and sense of self-efficacy and empowerment, with self-efficacy and empowerment boosting their sense of autonomy. Where improved empowerment mainly pertained to interactional and behavioral constructs [48,49], such as meaning-giving and being of value to peers through providing support, gains in self-efficacy mostly concerned intrapersonal constructs such as being informed about treatments. With respect to modes of user engagement, three styles were identified, starting with reading only, evolving into posting, and culminating into responding. Individually and together, these participation styles related differentially to the users' (overall positive) appreciation of the platform. As a truly interactive engagement style, *responding* played a key role in empowering users, and being valuable to others boosted their belief in their own abilities (personal strength). Primarily, the participants used the forum to explore and try (new) coping and social skills for later use in their real lives. The central drawback of Depression Connect use was that some users had become too involved in the community, getting overwhelmed by the continuous supply of posts and messages and their empathy for their fellow users. Finally, they noted that the Depression Connect community had provided them with an emotionally safe context to reach out to others in addition to their seeking or receiving face-to-face support and professional care.

Empowerment

One definition of empowerment in the context of this study reads "health care professionals collaborating with patients to help them acquire knowledge and resources" [46], which implies that it requires an inherently unequal relationship—one between knowledgeable health professionals and uninformed patients—to acquire knowledge and skills in managing a condition. Because of this paternalistic interpretation, the construct of empowerment is being criticized, as it contradicts the collaborative nature of the process [47]. Together with earlier positive findings on ISGs [14,18,19], our results suggest that offline and online peer communities for depression can be quite helpful for users to learn about and try new management and coping techniques. The reciprocal and "same-level" character of peer support defies the paternalistic notion of empowerment [47]. In terms of empowering patients, interactions with peers may even supplement professional care given that sharing experiential knowledge is not part of the therapeutic relationship.

Considering empowerment is a process rather than a mere outcome [46,49], we found that use of the Depression Connect platform specifically supported processes such as helping others [43] and meaning-giving. Peer contacts, and particularly sharing experiential knowledge to support others, fostered an external focus, consistent with the assumption that ISGs promote interactional empowerment. As an integral part of the process toward empowerment [46], we found that self-efficacy was mainly boosted by intrapersonal processes (ie, gaining personalized information on depression and coping skills) mirroring intrapersonal empowerment. Accordingly, we presume

XSL•FO
RenderX

that participating in ISGs helps advance both intrapersonal and interpersonal or interactional empowerment.

Findings in Context

Development and Variation in User Engagement

The benefits the Depression Connect users we interviewed derived from the forum are consistent with findings of other studies: informational and emotional support [11,12,14-17], social companionship [11,26], and empowerment [16,22]. Exploring which mechanisms drive ISGs and Depression Connect in particular, we compared styles of user engagement with the users' judgments. Although the three participation styles we identified (reading, posting, and responding) all had their own merits, the users' narratives revealed differential patterns in their online behavior. As alluded to in the introduction, previous ISG studies generally distinguished "lurkers" (ie, readers) and "posters," that is, users with fixed behavior patterns [43]. However, our results suggest that due to the cyclical and erratic nature of depression participation styles tend to evolve and fluctuate. According to most participants, the autonomy in choosing how they engaged in Depression Connect was a core advantage of online peer support, distinguishing it from other forms of offline peer support or formal care. When faced with (recurrent) depression, people often feel compelled to keep functioning well in daily life, being a good spouse, mother or father, employee, friend, or even patient [51]. When seeking support online, they do not feel this pressure and can let themselves be guided by their current needs. Whether they translate this behavior and positive experience to everyday life remains unknown.

Moreover, the development of and variations in participation styles over time contributed to user satisfaction. After a passive start, users gained more confidence from reading others' posts and responses and became more (inter)active themselves, making the shift from reading only to asking for help, sharing experiences, and finally helping others. Posting and responding brought gratification, boosting the way they thought about themselves, adding to their self-confidence, which Schwartz termed the "response shift effect in peer support" [65]. Nevertheless, future investigations should confirm whether accessible online communities like Depression Connect facilitate the transfer of learned skills to daily life.

Participation Styles and Perceived Value of Online Peer Support

In addition to the development of and flexibility in user engagement over time, our data suggest a direct association between participation styles and the perceived value of Depression Connect as an online community, which expands the findings on depression ISG research [19]. We found that the hypothesized relations between participation modes and ISG appreciation are similar to processes and associations observed in mental health care. Thus, the relationship between responders and enhanced empowerment resembles the benefits people derive from the so-called "helper role" [65] during group sessions or peer support meetings. The positive effects of helping others by responding to their narratives, such as feeling useful [66,67], promotes empowerment, as is also reflected by the growing (self-)confidence Depression Connect users reported when they began responding to peers. The observed association between posting and emotional growth or emotion regulation (ie, increasing self-knowledge through reflection on coping processes) echoes the role of expressive writing in reducing psychological distress [27,68]. By posting, simply another form of expressive writing, Depression Connect users found themselves learning to express and control their emotions better. In sum, our findings show that ISG members use passive, active or interactive styles of engagement to seek and derive different types of support from online peer communities, dependent on their personal needs over time.

Practical Implications of ISG Use

In their systematic review, Leamy et al [30] pose that in the context of recovery-oriented mental health care, coping with depression exceeds self-management and clinical recovery. They propose important themes for personal recovery, including connectedness and empowerment [30], which correspond to the main advantages mentioned by Depression Connect users in our study. Hence, we posit that participation in an ISG may facilitate and possibly accelerate recovery (ie, improved symptom management), with users finding their own paths. Importantly, we found that the Depression Connect platform was mainly used in addition to professional psychological or psychopharmacological care, experiences with which were exchanged, with peers offering participants different, experiential perspectives on (coping with) depression. Since ISGs offer its members a more holistic approach to their mental health issues and associated problems, health professionals may consider recommending them to (some of) their clients to complement ongoing therapy or as a form of informal follow-up care after therapy discontinuation. As a matter, of course, they are advised to inform themselves and their clients of the potential adverse events associated with online fora [69].

Limitations

Depression Connect users we interviewed may not be representative of all Depression Connect members; apart from the 2 negative cases, most participants were probably among the more frequent users because they were the more likely to come across the invitation for participation we posted. Furthermore, because the interviews were conducted during the COVID-19 pandemic when face-to-face contact was restricted, the importance of online types of support for depression increased, potentially causing the results to be biased in a positive direction.

The high accessibility (ie, free and easy of use) of the Depression Connect platform, the encouraging but nondirective role of its moderators, and its structural embedding in both a patient and mental health organization may have fostered social and interactive processes (eg, connectedness and support) that may not be representative of other ISGs that are less closely monitored [70]. Moreover, since Depression Connect is a Dutch-language forum and all participants were Dutch, we do not know whether our findings can be generalized to ISGs in other countries. It is possible that Dutch users attribute a greater value to (online) peer support because such services are not embedded in regular depression care in contrast to other

XSL•FO
RenderX

countries, such as Germany [71]. Finally, the benefits our participants claimed to derive from the use of Depression Connect largely reflect short-term gains, as the duration of their forum participation varied from 1.5 to 11 months at the time of data analysis.

Future Research

In a quantitative parallel study, we are in the process of evaluating the effects of Depression Connect use on empowerment (primary outcome measure) after 3 and 6 months. Further longitudinal research should be aimed at the longer-term beneficial and adverse effects of participation in ISGs.

A mixed-method effectiveness study should address the complexity and potential of peer support interventions. The method can yield rich and comprehensive data and thus provide a more holistic view on how people cope with depression. In this context, examining the perceived level of social support in daily life in relation to user statistics of online peer support services will be informative. Finally, a key challenge is to determine whether skills learned from peers in online networks also contribute to mental health recovery in the offline world [72].

Conclusions

Users of Depression Connect considered the online peer support community an accessible and valuable tool for learning to cope (better) with their depression. Seeking to understand the working mechanisms of ISGs, we found that the greater majority of the study participants benefited from the freedom and flexibility Depression Connect offered, allowing them to employ passive, active, and interactive styles of user engagement depending on their current mood and needs. Most found the forum, monitored by experienced peers, a safe environment to practice social and coping skills for later deployment in the offline world, supplementing formal and informal care. We found that besides promoting intrapersonal empowerment, Depression Connect also fostered interactional empowerment. Provided platforms are closely monitored and used to complement or follow-up formal care, and pending further investigations, we suggest that online peer support may be recommended as a safe context for exchanging knowledge and experiences on how to cope with depression and practice newly gained insights and skills.

Acknowledgments

We would like to thank the Radboudumc REshape Center of the Radboud University Medical Center for their close collaboration and valuable contributions in developing Depression Connect. Additionally, we thank DigitalePoli.nl for building and hosting the Depression Connect website and online platform. Most importantly, we thank all the users of Depression Connect for their commitment, with special thanks to the members who participated so generously in this study. Finally, we gratefully acknowledge the time and effort all participating professionals put in our research and independent researcher Denise te Paste for coding the data for analysis. The “The Power of Depression” research project and current study are funded by ZonMw, the Netherlands Organization for Health Research and Development. The funder is not involved in the design of the study nor in the collection, analysis, and interpretation of the data or the writing of the manuscript.

Authors' Contributions

JV, BF, and JS conceptualized the study. DS developed the study design, as well as collected, analyzed, and interpreted the data, and drafted the manuscript. DS had full access to all the study data and had the final responsibility for submitting the manuscript for publication. AD contributed to the collection and interpretation of the data, and to the writing of the current draft. JV, BF, JS, and JP participated in the critical review and revision of the manuscript. All authors provided intellectual content, reviewed, edited, and amended the manuscript. All authors gave their final approval for the current version to be published and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 245 KB - jmir_v23i7e25917_app1.pdf]

References

1. Pfeiffer PN, Heisler M, Piette JD, Rogers MAM, Valenstein M. Efficacy of peer support interventions for depression: a meta-analysis. Gen Hosp Psychiatry 2011;33(1):29-36 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.10.002] [Medline: 21353125]
2. Miyamoto Y, Sono T. Lessons from peer support among individuals with mental health difficulties: a review of the literature. Clin Pract Epidemiol Ment Health 2012;8:22-29 [FREE Full text] [doi: 10.2174/1745017901208010022] [Medline: 22563347]

XSL•FO
RenderX

3. Lloyd-Evans B, Mayo-Wilson E, Harrison B, Istead H, Brown E, Pilling S, et al. A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. BMC Psychiatry 2014 Feb 14;14:39 [FREE Full text] [doi: 10.1186/1471-244X-14-39] [Medline: 24528545]
4. Griffiths KM, Calear AL, Banfield M. Systematic review on internet support groups (ISGS) and depression (1): do ISGS reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
5. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman ATF. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr Scand 2010 Sep;122(3):184-191. [doi: 10.1111/j.1600-0447.2009.01519.x] [Medline: 20003092]
6. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
7. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. Aust N Z J Psychiatry 2006 Jan;40(1):51-54. [doi: 10.1111/j.1440-1614.2006.01741.x] [Medline: 16403038]
8. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol 2000 Feb;55(2):205-217. [Medline: 10717968]
9. Millard RW, Fintak PA. Use of the internet by patients with chronic illness. Disease Management and Health Outcomes 2002;10(3):187-194. [doi: 10.2165/00115677-200210030-00006]
10. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
11. Griffiths KM, Calear AL, Banfield M, Tam A. Systematic review on Internet Support Groups (ISGs) and depression (2): What is known about depression ISGs? J Med Internet Res 2009 Sep 30;11(3):e41 [FREE Full text] [doi: 10.2196/jmir.1303] [Medline: 19793718]
12. Evans M, Donelle L, Hume-Loveland L. Social support and online postpartum depression discussion groups: a content analysis. Patient Educ Couns 2012 Jun;87(3):405-410. [doi: 10.1016/j.pec.2011.09.011] [Medline: 22019021]
13. Feldhege J, Moessner M, Bauer S. Who says what? Content and participation characteristics in an online depression community. J Affect Disord 2020 Feb 15;263:521-527. [doi: 10.1016/j.jad.2019.11.007] [Medline: 31780138]
14. Griffiths KM, Reynolds J, Vassallo S. An online, moderated peer-to-peer support bulletin board for depression: user-perceived advantages and disadvantages. JMIR Ment Health 2015;2(2):e14 [FREE Full text] [doi: 10.2196/mental.4266] [Medline: 26543919]
15. Moore D, Ayers S, Drey N. A thematic analysis of stigma and disclosure for perinatal depression on an online forum. JMIR mental health. JMIR Ment Health 2016 May 19;3(2):e18 [FREE Full text] [doi: 10.2196/mental.5611] [Medline: 27197516]
16. Nimrod G. From knowledge to hope: Online depression communities. Int J Disabil Hum Dev 2012;11(1):23-30. [doi: 10.1515/ijdhd.2012.009]
17. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
18. Breuer L, Barker C. Online support groups for depression: benefits and barriers. SAGE Open 2015 Apr 08;5(2):215824401557493 [FREE Full text] [doi: 10.1177/2158244015574936]
19. Nimrod G. Online depression communities: members' interests and perceived benefits. Health Commun 2013;28(5):425-434. [doi: 10.1080/10410236.2012.691068] [Medline: 22809441]
20. Takahashi Y, Uchida C, Miyaki K, Sakai M, Shimbo T, Nakayama T. Potential benefits and harms of a peer support social network service on the internet for people with depressive tendencies: qualitative content analysis and social network analysis. J Med Internet Res 2009 Jul 23;11(3):e29 [FREE Full text] [doi: 10.2196/jmir.1142] [Medline: 19632979]
21. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89. [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]
22. Crisp D, Griffiths K, Mackinnon A, Bennett K, Christensen H. An online intervention for reducing depressive symptoms: secondary benefits for self-esteem, empowerment and quality of life. Psychiatry Res 2014 Apr 30;216(1):60-66. [doi: 10.1016/j.psychres.2014.01.041] [Medline: 24534125]
23. Crisp DA, Griffiths KM. Reducing depression through an online intervention: benefits from a user perspective. JMIR Ment Health 2016 Jan 08;3(1):e4 [FREE Full text] [doi: 10.2196/mental.4356] [Medline: 26747378]
24. Morris RR, Schueller SM, Picard RW. Efficacy of a Web-based, crowdsourced peer-to-peer cognitive reappraisal platform for depression: randomized controlled trial. J Med Internet Res 2015 Mar 30;17(3):e72 [FREE Full text] [doi: 10.2196/jmir.4167] [Medline: 25835472]
25. Tomasino KN, Lattie EG, Ho J, Palac HL, Kaiser SM, Mohr DC. Harnessing peer support in an online intervention for older adults with depression. Am J Geriatr Psychiatry 2017 Oct;25(10):1109-1119 [FREE Full text] [doi: 10.1016/j.jagp.2017.04.015] [Medline: 28571785]

XSL•FO
RenderX

26. Goodwin BC, Ford DE, Hsiung RC, Houston TK, Fogel J, Van Voorhees BW. First, do no harm: referring primary care patients with depression to an internet support group. Telemed J E Health 2018 Jan;24(1):37-44 [FREE Full text] [doi: 10.1089/tmj.2017.0004] [Medline: 28657881]
27. Dean J, Potts HW, Barker C. Direction to an internet support group compared with online expressive writing for people with depression and anxiety: a randomized trial. JMIR Ment Health 2016 May 17;3(2):e12 [FREE Full text] [doi: 10.2196/mental.5133] [Medline: 27189142]
28. Rollman BL, Herbeck Belnap B, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
29. Geramita EM, Herbeck Belnap B, Abebe KZ, Rothenberger SD, Rotondi AJ, Rollman BL. The association between increased levels of patient engagement with an internet support group and improved mental health outcomes at 6-month follow-up: post-hoc analyses from a randomized controlled trial. J Med Internet Res 2018 Jul 17;20(7):e10402 [FREE Full text] [doi: 10.2196/10402] [Medline: 30021711]
30. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry 2011 Dec;199(6):445-452 [FREE Full text] [doi: 10.1192/bjp.bp.110.083733] [Medline: 22130746]
31. Hensel JM, Shaw J, Ivers NM, Desveaux L, Vigod SN, Cohen A, et al. A web-based mental health platform for individuals seeking specialized mental health care services: multicenter pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 04;21(6):e10838 [FREE Full text] [doi: 10.2196/10838] [Medline: 31165710]
32. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 04;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
33. Carron-Arthur B, Cunningham JA, Griffiths KM. Describing the distribution of engagement in an internet support group by post frequency: a comparison of the 90-9-1 Principle and Zipf's Law. Internet Interventions 2014 Oct;1(4):165-168. [doi: 10.1016/j.invent.2014.09.003]
34. Sun N, Rau PP, Ma L. Understanding lurkers in online communities: a literature review. Comput Human Behav 2014 Sep;38:110-117. [doi: 10.1016/j.chb.2014.05.022]
35. Carron-Arthur B, Reynolds J, Bennett K, Bennett A, Griffiths KM. What's all the talk about? topic modelling in a mental health internet support group. BMC Psychiatry 2016 Oct 28;16(1):367 [FREE Full text] [doi: 10.1186/s12888-016-1073-5] [Medline: 27793131]
36. Carron-Arthur B, Ali K, Cunningham JA, Griffiths KM. From help-seekers to influential users: a systematic review of participation styles in online health communities. J Med Internet Res 2015 Dec 01;17(12):e271 [FREE Full text] [doi: 10.2196/jmir.4705] [Medline: 26627369]
37. Nimrod G. The membership life cycle in online support groups. International Journal of Communication 2012;6:23. [doi: 10.4135/9781483346427.n384]
38. Salem DA, Bogat GA, Reid C. Mutual help goes on-line. J Community Psychol 1997 Mar;25(2):189-207. [doi: 10.1002/(sici)1520-6629(199703)25:2<189::aid-jcop7>3.0.co;2-t]
39. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Comput Human Behav 2008 Sep;24(5):1867-1883. [doi: 10.1016/j.chb.2008.02.004]
40. Bellamy C, Schmutte T, Davidson L. An update on the growing evidence base for peer support. Mental Health and Social Inclusion 2017 Jun 12;21(3):161-167. [doi: 10.1108/mhsi-03-2017-0014]
41. Melling B, Houguet-Pincham T. Online peer support for individuals with depression: a summary of current research and future considerations. Psychiatr Rehabil J 2011;34(3):252-254. [doi: 10.2975/34.3.2011.252.254] [Medline: 21208865]
42. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
43. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
44. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
45. Petrovčič A, Petrič G. Differences in intrapersonal and interactional empowerment between lurkers and posters in health-related online support communities. Comput Human Behav 2014 May;34:39-48. [doi: 10.1016/j.chb.2014.01.008]
46. Cerezo PG, Juvé-Udina ME, Delgado-Hito P. Concepts and measures of patient empowerment: a comprehensive review. Rev Esc Enferm USP 2016;50(4):667-674 [FREE Full text] [doi: 10.1590/S0080-623420160000500018] [Medline: 27680054]
47. Halvorsen K, Dihle A, Hansen C, Nordhaug M, Jerpseth H, Tveiten S, et al. Empowerment in healthcare: A thematic synthesis and critical discussion of concept analyses of empowerment. Patient Educ Couns 2020 Jul;103(7):1263-1271 [FREE Full text] [doi: 10.1016/j.pec.2020.02.017] [Medline: 32164960]

XSL•FO
RenderX

48. Zimmerman MA. Psychological empowerment: issues and illustrations. Am J Community Psychol 1995 Oct;23(5):581-599. [doi: 10.1007/BF02506983] [Medline: 8851341]
49. Miguel MC, Ornelas JH, Maroco JP. Defining psychological empowerment construct: analysis of three empowerment scales. J Community Psychol 2015 Aug 12;43(7):900-919. [doi: 10.1002/jcop.21721]
50. Agner J, Braun KL. Patient empowerment: a critique of individualism and systematic review of patient perspectives. Patient Educ Couns 2018 Dec;101(12):2054-2064. [doi: 10.1016/j.pec.2018.07.026] [Medline: 30143254]
51. Smit D, Peelen J, Vrijsen JN, Spijker J. An exploration of the conditions for deploying self-management strategies: a qualitative study of experiential knowledge in depression. BMC Psychiatry 2020 May 11;20(1):210 [FREE Full text] [doi: 10.1186/s12888-020-02559-3] [Medline: 32393226]
52. Flavian C, Gurrea R, Orús C. Web design: a key factor for the website success. J of Systems and Info Tech 2009 May;11(2):168-184. [doi: 10.1108/13287260910955129]
53. Swain J. A hybrid approach to thematic analysis in qualitative research: Using a practical example. London: SAGE Research Methods Cases Part 2; 2018:1526435470.
54. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Glaser BG, Strauss AL. Discovery of grounded theory: Strategies for qualitative research. Abingdon: Routledge; 2017.
56. Strauss AL. Qualitative analysis for social scientists. Cambridge: Cambridge University Press; 1987.
57. Boeije H. Analysis in qualitative research. London: SAGE Publication Ltd; 2009.
58. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. American Journal of Evaluation 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
59. Maso I, Smaling A. Kwalitatief Onderzoek: Theorie en Praktijk [Qualitative research: theory and practice]. Book in German. Amsterdam: Boom; 1998.
60. Boeije H. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. Quality and quantity 2002;36(4):391-409. [doi: 10.4135/9781849209403.n920]
61. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. nt J Qual Methods 2020 Jan 22;19:160940691989922-160940691989913. [doi: 10.1177/1609406919899220]
62. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016 Nov;26(13):1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
63. Hanley T, Prescott J, Gomez KU. A systematic review exploring how young people use online forums for support around mental health issues. J Ment Health 2019 Oct;28(5):566-576. [doi: 10.1080/09638237.2019.1630725] [Medline: 31267813]
64. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
65. Schwartz CE, Sendor M. Helping others helps oneself: response shift effects in peer support. Soc Sci Med 1999 Jun;48(11):1563-1575. [doi: 10.1016/s0277-9536(99)00049-0] [Medline: 10400257]
66. Watson E. The mechanisms underpinning peer support: a literature review. J Ment Health 2019 Dec;28(6):677-688. [doi: 10.1080/09638237.2017.1417559] [Medline: 29260930]
67. Solomon P. Peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Rehabil J 2004;27(4):392-401. [doi: 10.2975/27.2004.392.401] [Medline: 15222150]
68. Frattaroli J. Experimental disclosure and its moderators: a meta-analysis. Psychol Bull 2006 Nov;132(6):823-865. [doi: 10.1037/0033-2909.132.6.823] [Medline: 17073523]
69. Easton K, Diggle J, Ruethi-Davis M, Holmes M, Byron-Parker D, Nuttall J, et al. Qualitative exploration of the potential for adverse events when using an online peer support network for mental health: cross-sectional survey. JMIR Ment Health 2017 Oct 30;4(4):e49 [FREE Full text] [doi: 10.2196/mental.8168] [Medline: 29084710]
70. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
71. Matzat J. The development of self-help groups and support for them in Germany. Int J Self Help Self Care 2001 Jan 1;1(4):307-322. [doi: 10.2190/9RVD-KPEC-M8PF-DDQN]
72. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]

Abbreviations

ISG: internet support group
MHISG: mental health internet support group
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by R Kukafka; submitted 20.11.20; peer-reviewed by S Watty, G Petrič, H Pratomo; comments to author 15.01.21; revised version received 22.04.21; accepted 24.05.21; published 12.07.21.

Please cite as:
Smit D, Vrijsen JN, Groeneweg B, Vellinga-Dings A, Peelen J, Spijker J
A Newly Developed Online Peer Support Community for Depression (Depression Connect): Qualitative Study
J Med Internet Res 2021;23(7):e25917
URL: https://www.jmir.org/2021/7/e25917
doi:10.2196/25917
PMID:34255659

©Dorien Smit, Janna N Vrijsen, Bart Groeneweg, Amber Vellinga-Dings, Janneke Peelen, Jan Spijker. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Support Seeking in the Postpartum Period: Content Analysis of Posts in Web-Based Parenting Discussion Groups

Bonnie R Chivers[1], MHumNutr; Rhonda M Garad[1], PhD; Lisa J Moran[1], PhD; Siew Lim[1], PhD; Cheryce L Harrison[1,2], PhD

[1]Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Clayton, Australia
[2]Diabetes and Vascular Medicine, Monash Health, Clayton, Australia

Corresponding Author:
Cheryce L Harrison, PhD
Monash Centre for Health Research and Implementation
School of Public Health and Preventive Medicine
Monash University
Level 1, 43-51 Kanooka Grove
Clayton, 3168
Australia
Phone: 61 3 8572 2600
Email: cheryce.harrison@monash.edu

Abstract

Background: The transition from pregnancy to motherhood is a major developmental phase that can be challenging for both women and their families. For new mothers, the postpartum period is recognized as a critical period for increased risk of both physical and mental health concerns. For this reason, it is imperative that women receive accurate, evidence-based information during this time.

Objective: This study aims to explore the conversations of new mothers on a web-based parenting forum to investigate what topics or concerns are being discussed.

Methods: A leading Australian web-based support forum for women before and after birth was used to obtain a sample of posts from the mothers of infants aged 0-12 months. Quantitative data (word frequencies and sentiment analysis) and qualitative data (post content) were extracted from discussion threads and examined to determine sentiments and theoretical storylines.

Results: In total, 260 posts were sampled. Infant care was the most prominent overarching topic discussed, with feeding and sleep being the most discussed subtopics. Discussions about maternal care were much less frequent but included questions about birth recovery, breastfeeding concerns, and interconception. A pattern of behavior emerged within the posts. This pattern resembled a cycle of learning across five phases: help seeking, solution ideation, testing and skill development, consolidation, and empowerment and improved mental well-being. A dynamic interplay was observed as mothers navigated new concerns or developmental changes.

Conclusions: Engagement in web-based forums to seek help and support during the postpartum period was common, with infant health and well-being being the primary concerns for new mothers during this time. The identification of a maternal learning cycle within the forum underscores the contributory role of web-based communities in maternal peer social support, information seeking, and early parenting practices.

(J Med Internet Res 2021;23(7):e26600) doi:10.2196/26600

KEYWORDS

pregnancy; perinatal; maternal; postpartum; infant; social support; qualitative; health; online

Introduction

The transition from pregnancy to motherhood is a major developmental phase that is recognized as a challenging time for both women and their families [1]. Early parenting and infant care are often prioritized over the health of the mother, presenting as significant barriers to self-care in the early postpartum period [2-4]. These barriers inhibit efforts to maintain or improve overall health, mental health, and healthy lifestyle behaviors [2-4], such as adequate diet quality and

XSL•FO
RenderX

regular physical activity [5]. For these reasons, the postpartum period is recognized as a critical period for an increased risk of adverse health. Weight retention after pregnancy is common [6,7] and is associated with excessive gestational weight gain [8], which is a strong predictor for the development of future obesity and chronic diseases [7]. Furthermore, mental health disorders, including anxiety and depression, affect up to 20% of women following pregnancy [9,10]. Postpartum anxiety and/or depression can exert significant effects on the health and well-being of mothers, their partners, and other children and can exert a negative impact on infant development [11]. Therefore, given the vulnerability to adverse physical and mental health, new mothers are a unique population with specific health needs that require increased support as well as accurate and trustworthy health information and care.

During the postpartum period, almost three-quarters (73%) of Australian parents with children aged less than 5 years use websites, blogs, and web-based forums to obtain information about infant or child health and parenting [12-14], with similar findings reported internationally [15]. Australian households have a considerably high internet use, with 97% of households with children having internet access [16]. Previous research evaluating the drivers of internet use during this time reported convenience, anonymity, and social and peer support as facilitating factors [13,17,18]. Web-based parenting forums are a common platform in which women can connect with peers for emotional support; alleviate feelings of isolation; and facilitate the discussion of sensitive topics that are otherwise difficult to address with friends, family, or health care providers in face-to-face encounters [18]. Yet, although reasons for engagement are clear [1,12,13,18], there is limited evidence showing what information and support women seek during the postpartum period and how they interact within such forums. An improved understanding of the information and support needs of women during this significant life phase is crucial to ensure that health and information needs of the mothers are met.

To address this research gap, this paper examines the unmediated user-generated content from web-based forum discussions of women in the postpartum period to identify early parenting information and emotional support–seeking behaviors.

Methods

Overview

An observational analysis of web-based discussions was conducted within a leading Australian internet discussion forum for new or expecting parents. The most popular Australian pre- and postbirth forum was identified by searching the term *new mum forum* on Google. The top 10 (first page) results were assessed, and all the websites with publicly available forums (n=7) were analyzed using a website analytics tool (Alexa, Amazon). This software was used to determine the global page views, global rank, and Australian rank of the seven websites with publicly available forums. The highest ranked website for Australian users was identified and used as the sampling platform for this study. To confirm the suitability of this website for this study, member requirements were assessed to ensure that forum users were new or expecting mothers. The second highest ranked site did not have an open access forum, and those ranked 3-10 had a significantly lower rank than the first and second highest ranked sites. Therefore, the first site was retained, with the remainder deemed insufficient for sampling. This approach is comparable with previous research that used discussion forums [19,20].

The selected forum allows members to interact with their *birth club*, which corresponds to their child's due date (month and year). Birth clubs are a subforum of the wider forum community, and they are nonspecific in the nature of discussion topics. In total, 13 birth clubs (January 2019 to January 2020) were selected as the sampling platform to represent 1 calendar year and therefore one cross-section of the postpartum period across this time (ie, <1 month to 12 months postpartum).

Included posts were sampled by selecting the first 20 posts or threads from each birth club at the time of collection. Posts were collected between January 6, 2020, and January 13, 2020. The exclusion criteria included posts that enquired about or discussed an elder sibling (not the infant aged 0-12 months), other people's child or children, extended family such as grandparents or in-laws, or products or shopping (unless the post also discussed infant care such as feeding product advice). Posts in the January 2020 birth club were excluded if the forum user indicated that they had not yet given birth. Posts were collected sequentially, as they appeared on the day of sampling, and if a post met the exclusion criteria, the following post was selected until 20 posts were obtained. Posts were extracted in a deidentified format into an Excel (Microsoft) document that included the post title, date, and content, comparable with previous research [19,20].

Analysis

Data were processed using NVivo Pro 12 software (QSR International) [21]. A modified grounded theory analysis was conducted, which was informed by the six-phase approach by Braun and Clarke [22]. Due to the understudied nature of parenting forums, a grounded theory approach is well suited to add depth and breadth to this investigation [23]. A single researcher (BRC) generated initial codes and then grouped them into core categories. Three authors agreed on initial and intermediate codes and conducted a narrative overview of the discussions (BRC, RMG, and CLH). A >10% (26/260) check was conducted after initial coding and during theme conceptualization by 2 additional researchers (RMG and CLH). As the themes were conceptualized, the research team developed a theoretical storyline, which could be observed beyond *what* was being discussed.

To support these findings, NVivo Pro 12 text frequency search was used to identify the common terms, thereby identifying prominent conversation topics. Word frequency calculations identified all stemmed words (minimum three letters). NVivo Pro 12 automatic sentiment analysis was performed to identify the emotional indicators. NVivo searched for the expressions of sentiment in the source material (forum posts) and used predefined scores for words classified as containing sentiment [24]. Words are considered in isolation, and the program then determines the sentiment of the paragraph as a calculation of each word containing the sentiment. Sentiment results include the number of references (paragraphs with sentiment) that are

XSL•FO
RenderX

categorized as very positive, moderately positive, moderately negative, and very negative [24].

Ethics

The Monash Health (RES-19-0000-291A) and Monash University (project no. 20196) Human Research Ethics Committees granted ethics approval for this study. Although the ethical oversight of publicly available data is not strictly required, the authors sought approval as per the Monash University protocol.

Results

Overview

In total, 260 posts were extracted and analyzed. The 13 birth clubs had an average of 3013 members in each club (n=39,163 forum members overall).

Analysis of User-Generated Content

The analysis of posts through open coding identified 432 references at the intermediate coding stage. Various posts discussed multiple topics; therefore, the number of topic references exceeded the number of posts. Infant-focused references were the most frequent (237/432, 54.8%), with 12.5% (54/432) of references relating to sleep and naps. References to infant health (46/432, 10.6%) and feeding were frequent (68/432, 15.7%), and 7.6% (33/432) of references were related to breastfeeding. Forum use to seek help, support, advice, or reassurance was frequent (71/432, 16.4%). Discussion topics relating to infant care commonly centered on health (eg, nappy rash, cracked lips, and cradle cap) and development (eg, common milestones including teething and sitting, crawling, or walking). Both first time and mothers with older children were active in these discussions. Women regularly used the forum to ease concerns and to assist them in times of need or confusion during their first year of motherhood.

Maternal health needs and/or well-being were less frequent, with the overall identification of 21.1% (91/432) of references. Most maternal health discussions were observed in the early postpartum period and became less frequent further on from the birth experience. Topics pertaining to maternal health included birth recovery, breastfeeding difficulties, mastitis or breast discomfort, pelvic floor health, and resumption of menstruation. There was limited discussion about modifiable health factors, including the mother's weight, exercise, or diet. The evidence of mental distress was observed with some women discussing the feelings of anxiety, birth trauma, or unhappiness (Table 1).

The discussion forum was used in tandem with care or advice from health professionals, not in place of it. Women appeared to use the forum to confirm a health issue, seek out the experiences of other mothers, or share their experiences. There was no evidence of disregarding the health advice from the health professionals.

Mothers reached out to other forum users when they were unsure about how to manage something and sought a similar experience from others in their birth club commonly asking the following: "Has anyone else had this?," "Has anyone else been told this?," "Can anyone else relate?," and "Anyone going through the same thing?" These inquiries match efforts to normalize experiences or to confirm a problem. Mothers were often observed describing a problem to ask others if this was *normal* and to determine if they should seek advice from health professionals: "Should I be concerned?," "What do I do? Should I take him to the doctor?...or is this normal?," and "Is this something I should be worried about at this age." Some mothers used the forum to allay worries as they bridged time until they could reach their doctor: "I am taking him to the doctors tomorrow but I just wanted to know if anyone has experienced this."

XSL•FO
RenderX

Table 1. Intermediate coding references (N=432).

Topic, subthemes, and references	Codes, n (%)
Infant	237 (54.9)
Sleep	54 (12.5)
Daytime or nighttime sleep routine	30 (6.9)
Bad sleeper	5 (1.1)
Hunger and sleep relationship	4 (0.9)
Cosleeping	3 (0.7)
Sleep regression or changes	4 (0.9)
Sleep training	3 (0.7)
Clothes or swaddle for sleep	2 (0.5)
Not sleeping due to teething	2 (0.5)
Sleep safety	1 (0.2)
Infant health	46 (10.6)
Skin concerns or topical treatments	15 (3.5)
Miscellaneous	12 (2.7)
Weight concerns	4 (0.9)
Bowel movements	4 (0.9)
Immunizations	4 (0.9)
Blood or mucus in nappy	3 (0.7)
Common cold	2 (0.5)
Tongue tie	2 (0.5)
Routines	39 (9)
Daytime nap routine	15 (3.5)
Nighttime sleep routine	15 (3.5)
Feeding routines	9 (2.1)
Feeding	35 (8.1)
Feeding solids	12 (2.7)
Formula amount or recommendations	6 (1.4)
Feeding routines	6 (1.4)
Refusing bottle	3 (0.7)
Unusual food-related behavior	3 (0.7)
Feeding cow’s milk	2 (0.5)
Unable to burp or upset tummy	2 (0.5)
Dad wanting to help feed	1 (0.2)
Breastfeeding	33 (7.6)
Milk supply	20 (4.6)
Breastfeeding routines	3 (0.7)
Pain or discomfort	3 (0.7)
Breastfeeding in subsequent pregnancy	2 (0.5)
Weaning	2 (0.5)
Drinking alcohol and breastfeeding	1 (0.2)
Feeding aides, that is, shields	1 (0.2)
Number of breastfeeds	1 (0.2)

XSL•FO
RenderX

Topic, subthemes, and references	Codes, n (%)
Development	19 (4.4)
Teething	7 (1.6)
Leap	5 (1.1)
Crawling	3 (0.7)
Talking	2 (0.5)
Walking	2 (0.5)
Infant behavior	6 (1.4)
Behavioral problems or concerns	6 (1.4)
Miscellaneous	5 (1.2)
Haircuts or ear piercing	2 (0.5)
Travel	2 (0.5)
Car seats	1 (0.2)
Maternal	156 (36.1)
Help seeking	71 (16.4)
Seeking emotional support	15 (3.4)
Seeking advice or reassurance	56 (12.9)
Psychosocial health	30 (6.9)
Emotional well-being	17 (3.9)
Anxious about something	4 (0.9)
Feeling lost or guilty	4 (0.9)
Lack of support	3 (0.7)
Changes and challenges	2 (0.5)
Interconception	25 (5.8)
Sleep deprived	6 (1.4)
Menstrual cycle returning	6 (1.4)
Subsequent pregnancy announcement	6 (1.4)
Becoming pregnant again (views or concerns)	4 (0.9)
Birth control	2 (0.5)
Irregular periods	1 (0.2)
Birth recovery or physical health	24 (5.6)
Natural birth recovery	12 (2.8)
C-section recovery	4 (0.9)
Weight loss	4 (0.9)
Birth experience	2 (0.5)
Stretch marks	1 (0.2)
Pelvic floor	1 (0.2)
Socializing	6 (1.4)
Networking	6 (1.4)
Medical	39 (9)
Health provider advice	23 (5.3)
Querying the advice of health provider (general physician or maternal child health nurse) with forum community	13 (3)
Discussing concern before seeking professional advice	10 (2.3)

XSL•FO
RenderX

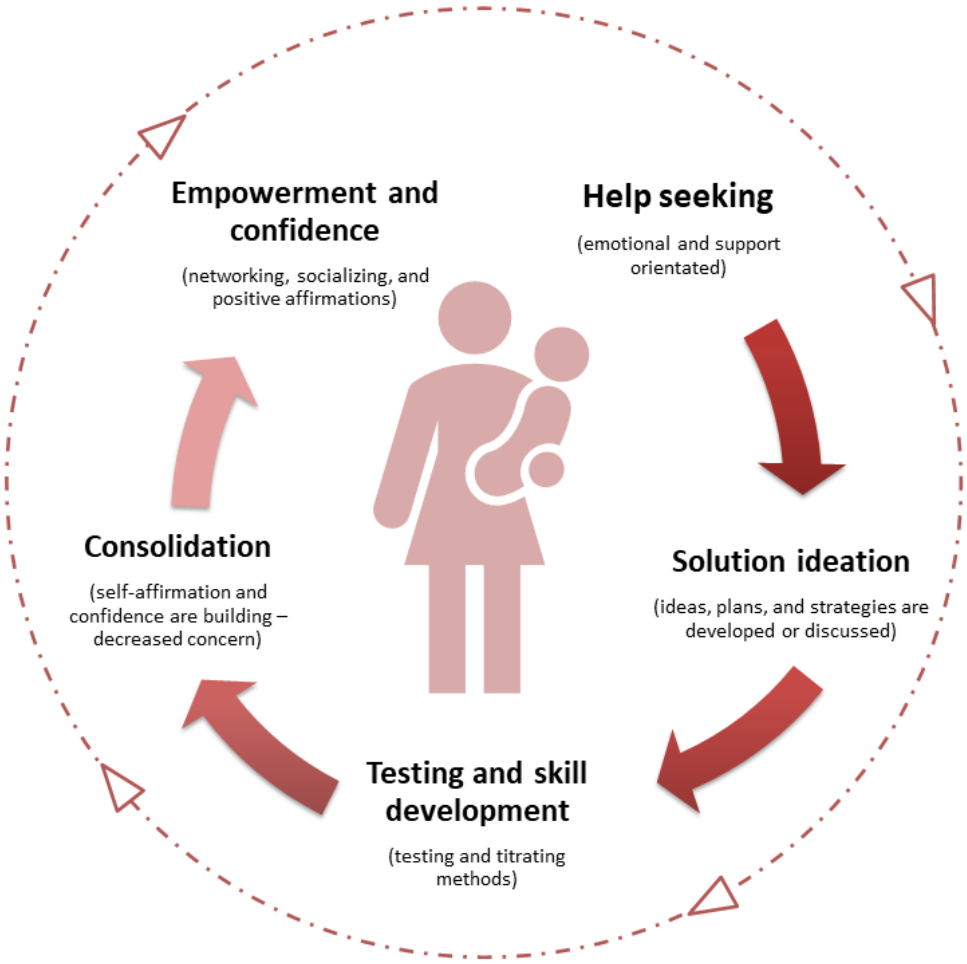
Topic, subthemes, and references	Codes, n (%)
Questioning	16 (3.7)
Clinically relevant questions	16 (3.7)

Overview of Advanced Coding

An emerging cross-cutting pattern through posts was observed while examining web-based discussions in the cohort. Although the discussion topics identified remained consistent, the mother's approach to discussing these concerns, how the concern was expressed, and level of support sought varied. This pattern was reflective of a cycle of learning across five phases: (1) help seeking, (2) solution ideation, (3) testing and skill development, (4) consolidation, and (5) empowerment and improved mental well-being (Figure 1). As mothers moved through the stages of parenting and learning, confidence among forum members increased and confusion, uncertainty, and the need for reassurance decreased. During the final stages of the learning process, women often presented as more empowered and self-assured, with some sharing stories of reassurance and others assuming the role of peer support providers to mothers in the early postpartum period. Although we describe a process model, in practice, we believe that early parenting experiences and parental skill development by mothers are characterized through a dynamic interplay within the model. Mothers may interact with different phases simultaneously, as they encounter new challenges that cause stress and uncertainty; yet, they are further progressed and more confident in other aspects of early parenting. For example, a mother may feel confident about infant feeding (later phases) and yet be anxious and uncertain about how to manage teething (earlier phases).

There appeared to be a temporal relationship between the length of time parenting and increased confidence. This is presumed and cannot be confirmed by this study; however, early postpartum mothers expressed greater levels of uncertainty and concern than those interacting with the forum during the later postpartum period. Overall, the vast majority of posts were from women expressing phases one or two, women who portrayed distress and uncertainty, and women who were more likely to seek support and advice from others. A minority of posts belonged to the later phases of the learning cycle, indicating that women were potentially more reassured and less inclined to engage with the forum specifically to share positive experiences.

Figure 1. Schematic of the learning cycle.



XSL•FO
RenderX

Five-Phase Learning Process (Drawn From Advanced Coding)

Phase One: Help Seeking

Women using the forum initiated discussions on an issue, need, or problem relating to their child or children or parental experiences. Conveying emotional experiences was common in this phase, as was maternal uncertainty. Women regularly stated that they were feeling *anxious*, *stressed*, *worried*, or *exhausted*. The forum acted as an outlet for these emotions and an opportunity to receive social and emotional support from other women:

> *our little miss is the worst sleeper ever [...] I am up and down all night we are lucky to get 3 hours [...] please any book ideas, throw them at me. I'm physically, emotionally and mentally drained.*
>
> *I'm really down at the moment [...] My 8-month-old doesn't sleep through, she doesn't show me any affection and is quite sooky and demanding. I [...] just feel like I'm waiting for the stage to pass so I can be happier and feel some sort of motivation.*
>
> *my little one is 9 weeks old and I still feel so clueless, does anyone else? [...] I just feel a bit lost.*

In the early postbirth period, many women used this stage to discuss their experiences before seeking health professional's advice or while bridging the time until they can seek help:

> *[I have the symptoms of an] episiotomy hematoma. Has anyone had any experiences with this? I'm checking in with my OBGYN tomorrow.*

Phase Two: Solution Ideation

Following the initial requests for help and/or information, women discussed potential solutions. Discussing or testing strategies or ways to resolve their concerns with their peers in the forum was common to this phase:

> *I need your help. My poor little baby is super constipated [...] Nothing has worked! I have tried pear and prunes, water, pear juice, Coloxyl, brown sugar and water.*

Solution ideation was also used by women building confidence to make changes:

> *for those of you who have started snacks for baby. What are you offering? Need ideas. Also breastfeeding mamas do you still offer a boob [breast] feed before putting baby down for a nap?*

Phase Three: Testing and Skill Development

In this phase, women started to implement strategies and test solutions. Using a trial-by-error approach, mothers titrated methods to obtain the best outcomes related to their concern. They discussed their results with their peers while seeking reassurance and guidance during this process:

> *recently increased my 16-week-old to 150ml and 5 bottles [...] she's struggling to take even 90ml at a time?! I [have] stretched to 4 hours thinking she might not be hungry but still no difference working. [...] she missing out on around 300 of the total, should I be worried? she's seems her usual happy self maybe napping a little more.*

During this phase, women have the confidence to try things or rationalize their experiences, yet require reassurance from their peers:

> *...is that too ambitious even for an 8-month-old? Does anyone else's baby not babble at all at 8 months? and/or what sounds are your babies making by now?*

Mothers often seek insight from the past experiences of others or their peers with a child of comparable age.

Phase Four: Consolidation

During this phase, women consolidated their new skills and practices. This phase was often coupled with an increased confidence and decreased uncertainty. It was common for women to post step-by-step outlines of their daily routines to compare with those at a similar stage. Women who anticipated that they were approaching a successful outcome were seen to reaffirm what they had tried or achieved, such as:

> *breastfeeding has never been easy, [...] however I was told to just persist. [...] He is already feeding a lot better than he was previously. Fingers crossed it gets better and better so I can go back to exclusively breastfeeding.*

A common process observed during this phase was the consolidation of advice received from health professionals:

> *she said that from 9 months, milk is secondary and food is to be offered first always. Has everyone else been told this? I'll follow her advice. Was just checking.*

Phase Five: Empowerment

In this final phase, women displayed a degree of empowerment characterized by an increased confidence in the use of acquired knowledge or skills. Many assumed the role of information provider to other mothers, which could be viewed via the responses to original posts, characterizing somewhat of a team working together to share ideas and support those in need. Within the original posts, we viewed women at this stage reaching out for connections, such as "how are all the mummas doing? just checking in," or networking "any mums living close to [...] who would like to connect." Those with the confidence to do so shared their experiences to support and guide others:

> *I wanted to create a thread in case you're feeling a bit down and want some encouragement from other women navigating their first/second/third/tenth time through the postpartum recovery journey. [...] It's not pretty, let's say that. [...] I'm tired but [...] I'm wandering around like an elderly lady, blissfully happy with our third born child [...] and feeling the pains, irks and exhaustion.*

Sentiment and Word Frequency Analysis

A word frequency calculation supports the findings of our open coding with *baby*, *sleep*, *feeds*, three of the five most frequently used words (Multimedia Appendix 1). There were 335

XSL•FO
RenderX

references of sentiment. Of these, most were found to be within the negative range, with 64.2% (215/335) of references classified as negative (very negative: n=126 and moderately negative: n=89) compared with 120 positive references (very positive: n=48 and moderately positive: n=78). A cross-check of sentiment results revealed that less than <10.1% (34/335) of the sentiment references were coded as incorrect categories.

Discussion

Principal Findings

This observational study examined unmediated peer-to-peer web-based discussions during the postpartum period, providing important insights into the information- and help-seeking needs of new mothers. We used the largest digital platform for new parents in Australia, which was representative of approximately 9.8% (30,000/305,832) of women giving birth annually [25], in line with engagement to opt in survey-based methods evaluating health behaviors [26]. Our results demonstrate a predominant focus on infant health needs, including feeding, breastfeeding, and sleep, during early parenting, with maternal health and well-being being a minor focus. Sentiment analyses revealed that the posts were more likely to be negatively portrayed, supporting the finding that the forum is commonly used to express a problem, seek information or help, and gain support or reassurance, consistent with previous literature [20]. The thematic analysis of posts revealed a pattern of behavior resembling a learning process whereby topics remained consistent, but how the concern was expressed and the level of support required varied. This process revealed several phases that commenced with help seeking through consolidation and empowerment. Overall, our findings provide critical insight into the concerns of new mothers and underscore the contributory role of web-based communities in maternal peer social support, information seeking, and the development of early parenting practices.

We report a central focus around the care and development of the infant, with a minority of posts about maternal well-being and fewer again centered on preventive health, including weight gain prevention or modifiable lifestyle factors. This finding is significant, given that the risk of adverse health is high, suboptimal lifestyle behaviors and weight gain are common, and barriers to health optimization exist during the postpartum period [2-4]. Previous studies have reported that time constraints are the most prominent barrier to healthy lifestyle changes, including physical activity engagement, at both 3 and 12 months after pregnancy [3]. In addition to reaffirming these barriers, our results provide additional insights and findings that personal health and well-being were not prominently discussed by mothers, suggesting that this is not a central priority during early parenting compared with that of infant health. This is potentially reflective of reduced engagement and compliance in postpartum healthy lifestyle interventions, as reported previously [27]. Subsequently, there is a paucity of effective strategies to engage women during this life phase for optimized health. Taken together, this emphasizes the need for novel approaches to enhance the awareness of, and engagement in, maternal preventive health during this period. This could potentially include maternal healthy lifestyle promotion delivered alongside infant care or design of holistic lifestyle programs including infants and wider family members to improve feasibility and engagement for new mothers. Alternatively, the implementation of healthy lifestyle programs during pregnancy when women are regularly engaged with health care providers has been shown to increase compliance in the postpartum period [28] and may better optimize lifestyle behaviors if practiced and maintained before birth.

Our results identified that discussion themes were underpinned by a learning development process, not dissimilar to those previously described, such as Becoming a Mother and Maternal Role Attainment theories [29-31]. The findings of this study emulate the concepts outlined in previous theories regarding maternal development (psychological adjustments and acquisition of a new role; acquaintance, learning, and physical restoration, which are both assumed and learned; and internalization of the maternal role, competence, and confidence). The crucial developmental processes at play within the forum emphasize the importance of the internet during this period and illustrate the influence forum communities may have on maternal decision-making and experiences. Lupton [18] previously identified that women use forums and web-based social networks to connect with other women and to gain guidance and insight through others' experiences and knowledge. We note that the majority of posts were in the early phases of the learning cycle, in which mothers were uncertain, requiring an increased emotional support or solution seeking from their peers. This is also reflected in the sentiment analysis results, with higher proportions of negative sentiment compared with positive sentiment. Our results show that an increasing maternal confidence potentially coincides with the skill and knowledge acquisition sought within web-based communities. Although not all knowledge and skills may be obtained through the forum, and the de-escalation of stress and uncertainty may be influenced by many factors, the forum community is clearly an important platform during early parenting knowledge acquisition for many women and, therefore, plays a significant role in the postpartum journey for new mothers.

Supporting parents to meet the challenges of their caregiving role has consistently been identified as a public health priority [32,33]. Despite this widespread recognition [34], knowledge gaps still exist regarding effective ways to promote positive parenting practices, and little evidence is available that clearly depicts how parents learn and develop [32]. Ensuring health professionals are aware of the support requirements of women during this phase as well as their information priorities, as identified here, is essential. Furthermore, the understanding that acquisition of knowledge and skills during early parenting is fluid and follows a learning cycle is important in enabling the provision of individualized information and support. Similarly, assisting women in recognizing these learning processes may alleviate postpartum stress [35] and anxiety experienced during the initial phases and, in turn, facilitate more rapid progression toward knowledge acquisition and confidence and possible improvements in mental well-being.

To our knowledge, this is the first study to assess the unmediated web-based conversations of mothers during the first year of the

XSL•FO
RenderX

postpartum period. Our findings portray parental experiences and perceptions without the influence of researchers or controlled research settings [36]. This design enables insight into the communication and output of emotions that women may experience at any time of the day, which may be lost to recall or have subsided when sought in a research or clinical setting.

Limitations

The following limitations should be considered while interpreting these findings. Anonymized data were interrogated, and therefore, demographic and geographical user information could not be obtained. User demographics may influence engagement with the forum as well as postpartum needs and experiences. Although users cannot be demographically profiled, the anonymous nature of the forum allows for uninhibited discussion, providing rich data on the needs of participating women. Due to site management and restrictions relating to the seeking or provision of medical advice via the forum, posts may have been deleted before data were collected.

Conclusions

The postpartum period involves a major life transition that requires increased levels of social, emotional, and health professional support. Our results demonstrate that engagement in web-based forums to seek help and support during the postpartum period is common, with infant health and well-being being primary concerns for new mothers during this time. A lack of discussion about maternal health was observed, emphasizing the need for improved awareness and novel engagement strategies. The identification of a maternal learning cycle at play within the forum demonstrates the significant role of web-based communities in maternal social support and in defining parenting. Further exploration is needed to understand how health care professionals can provide targeted and personalized support to women in postpartum period, where infant needs are prioritized above their own, particularly for those experiencing increased levels of distress and uncertainty.

Acknowledgments

We acknowledge Medibank for providing funding support for this research. BRC is supported by a RTP Postgraduate scholarship. SL and CLH are supported by the National Health and Medical Research Council (NHMRC) via an Early Career Fellowship (GNTAPP1139481) and a CRE Health in Preconception and Pregnancy Senior Postdoctoral Fellowship (APP1171142), respectively. LJM is funded by a Heart Foundation Future Leader Fellowship.

Authors' Contributions

Conceptulization and study design: CLH, BRC, RMG. Data extraction, collation, and analysis: BRC. Thematic analysis and interpretation: CLH, BRC, RMG. Manuscript preparation: BRC. All authors reviewed the manuscript for intellectual content and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequently used words in forum user-generated content (word cloud).
[PNG File , 77 KB - jmir_v23i7e26600_app1.png]

References

1. Lupton D. 'A love/hate relationship': the ideals and experiences of first- time mothers. J Sociol 2016 Jun 30;36(1):50-63. [doi: 10.1177/144078330003600104]
2. Graham M, Uesugi K, Olson C. Barriers to weight-related health behaviours: a qualitative comparison of the socioecological conditions between pregnant and post-partum low-income women. Matern Child Nutr 2016 Apr 18;12(2):349-361 [FREE Full text] [doi: 10.1111/mcn.12135] [Medline: 25040706]
3. Evenson KR, Aytur SA, Borodulin K. Physical activity beliefs, barriers, and enablers among postpartum women. J Womens Health (Larchmt) 2009 Dec;18(12):1925-1934 [FREE Full text] [doi: 10.1089/jwh.2008.1309] [Medline: 20044854]
4. Carter-Edwards L, Østbye T, Bastian LA, Yarnall KS, Krause KM, Simmons T. Barriers to adopting a healthy lifestyle: insight from postpartum women. BMC Res Notes 2009 Aug 17;2(1):161 [FREE Full text] [doi: 10.1186/1756-0500-2-161] [Medline: 19686601]
5. Martin J, Moran L, Teede H, Ranasinha S, Lombard C, Harrison C. Diet quality in a weight gain prevention trial of reproductive aged women: a secondary analysis of a cluster randomized controlled trial. Nutrients 2018 Dec 27;11(1):49 [FREE Full text] [doi: 10.3390/nu11010049] [Medline: 30591672]
6. Rong K, Yu K, Han X, Szeto IM, Qin X, Wang J, et al. Pre-pregnancy BMI, gestational weight gain and postpartum weight retention: a meta-analysis of observational studies. Public Health Nutr 2015 Aug;18(12):2172-2182. [doi: 10.1017/S1368980014002523] [Medline: 25411780]

XSL•FO
RenderX

7. Rooney B, Schauberger C. Excess pregnancy weight gain and long-term obesity: one decade later. Obstet Gynecol 2002 Aug;100(2):245-252. [doi: 10.1016/s0029-7844(02)02125-7] [Medline: 12151145]
8. Harrison CL, Lombard CB, Teede HJ. Limiting postpartum weight retention through early antenatal intervention: the HeLP-her randomised controlled trial. Int J Behav Nutr Phys Act 2014 Oct 31;11(1):134-138 [FREE Full text] [doi: 10.1186/s12966-014-0134-8] [Medline: 25358909]
9. Fawcett EJ, Fairbrother N, Cox ML, White IR, Fawcett JM. The Prevalence of Anxiety Disorders During Pregnancy and the Postpartum Period: A Multivariate Bayesian Meta-Analysis. J Clin Psychiatry 2019 Jul 23;80(4):18r12527 [FREE Full text] [doi: 10.4088/JCP.18r12527] [Medline: 31347796]
10. Falah-Hassani K, Shiri R, Dennis C. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med 2017 Sep;47(12):2041-2053. [doi: 10.1017/S0033291717000617] [Medline: 28414017]
11. Rogers A, Obst S, Teague SJ, Rossen L, Spry EA, Macdonald JA, et al. Association between maternal perinatal depression and anxiety and child and adolescent development: a meta-analysis. JAMA Pediatr 2020 Nov 1;174(11):1082-1092. [doi: 10.1001/jamapediatrics.2020.2910] [Medline: 32926075]
12. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171-110 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
13. Lupton D, Maslen S. How women use digital technologies for health: qualitative interview and focus group study. J Med Internet Res 2019 Jan 25;21(1):e11481 [FREE Full text] [doi: 10.2196/11481] [Medline: 30681963]
14. Sources of Child Health Information: What Parents Use and Trust. Royal Children's Hospital National Child Health Poll. 2016. URL: https://www.rchpoll.org.au/polls/sources-of-child-health-information-what-parents-use-and-trust/ [accessed 2021-07-02]
15. Goldman RD, Macpherson A. Internet health information use and e-mail access by parents attending a paediatric emergency department. Emerg Med J 2006 May 1;23(5):345-348 [FREE Full text] [doi: 10.1136/emj.2005.026872] [Medline: 16627833]
16. Household Use of Information Technology. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/latest-release [accessed 2020-12-18]
17. Laws R, Walsh AD, Hesketh KD, Downing KL, Kuswara K, Campbell KJ. Differences between mothers and fathers of young children in their use of the internet to support healthy family lifestyle behaviors: cross-sectional study. J Med Internet Res 2019 Jan 23;21(1):e11454 [FREE Full text] [doi: 10.2196/11454] [Medline: 30674450]
18. Lupton D. 'it just gives me a bit of peace of mind': Australian women's use of digital media for pregnancy and early motherhood. Societies 2017 Sep 15;7(3):25. [doi: 10.3390/soc7030025]
19. Chivers BR, Garad RM, Boyle JA, Skouteris H, Teede HJ, Harrison CL. Perinatal distress during covid-19: thematic analysis of an online parenting forum. J Med Internet Res 2020 Sep 7;22(9):e22002 [FREE Full text] [doi: 10.2196/22002] [Medline: 32857707]
20. Ellis L, Roberts L. Exploring the use and quality of Internet discussion forums in pregnancy: a qualitative analysis. Birth 2020 Mar 4;47(1):153-161. [doi: 10.1111/birt.12459] [Medline: 31583769]
21. NVivo Qualitative Data Analysis Software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-07-06]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Charmaz K, Belgrave L. Grounded theory. In: The Blackwell Encyclopedia of Sociology. Oxford, UK: Blackwell Publishing Ltd; 2007.
24. How Auto Coding Sentiment Works. QSR International. URL: http://help-nv11.qsrinternational.com/desktop/concepts/How_auto_coding_sentiment_works.htm [accessed 2021-02-07]
25. Births, Australia. Australian Bureau of Statistics. 2019. URL: https://www.abs.gov.au/statistics/people/population/births-australia/latest-release [accessed 2021-02-21]
26. Chivers BR, Boyle JA, Lang AY, Teede HJ, Moran LJ, Harrison CL. Preconception health and lifestyle behaviours of women planning a pregnancy: a cross-sectional study. J Clin Med 2020 Jun 2;9(6):1701 [FREE Full text] [doi: 10.3390/jcm9061701] [Medline: 32498329]
27. Haste A, Adamson AJ, McColl E, Araujo-Soares V, Bell R. Problems recruiting and retaining postnatal women to a pilot randomised controlled trial of a web-delivered weight loss intervention. BMC Res Notes 2018 Mar 27;11(1):203-207 [FREE Full text] [doi: 10.1186/s13104-018-3305-x] [Medline: 29587868]
28. Ferrara A, Hedderson MM, Albright CL, Ehrlich SF, Quesenberry CP, Peng T, et al. A pregnancy and postpartum lifestyle intervention in women with gestational diabetes mellitus reduces diabetes risk factors: a feasibility randomized control trial. Diabetes Care 2011 Jul 3;34(7):1519-1525 [FREE Full text] [doi: 10.2337/dc10-2221] [Medline: 21540430]
29. Barclay L, Everitt L, Rogan F, Schmied V, Wyllie A. Becoming a mother--an analysis of women's experience of early motherhood. J Adv Nurs 1997 Apr;25(4):719-728. [doi: 10.1046/j.1365-2648.1997.t01-1-1997025719.x] [Medline: 9104667]
30. Mercer RT. Becoming a mother versus maternal role attainment. J Nurs Scholarsh 2004 Sep;36(3):226-232. [doi: 10.1111/j.1547-5069.2004.04042.x] [Medline: 15495491]

XSL•FO
RenderX

31. Rogan F, Shmied V, Barclay L, Everitt L, Wyllie A. "Becoming a mother'--developing a new theory of early motherhood. J Adv Nurs 1997 May 28;25(5):877-885. [doi: 10.1046/j.1365-2648.1997.1997025877.x] [Medline: 9147193]
32. Sanders MR. Triple p-positive parenting program as a public health approach to strengthening parenting. J Fam Psychol 2008 Aug;22(4):506-517. [doi: 10.1037/0893-3200.22.3.506] [Medline: 18729665]
33. Butler J, Gregg L, Calam R, Wittkowski A. Parents' perceptions and experiences of parenting programmes: a systematic review and metasynthesis of the qualitative literature. Clin Child Fam Psychol Rev 2020 Jun 10;23(2):176-204 [FREE Full text] [doi: 10.1007/s10567-019-00307-y] [Medline: 31820298]
34. Britto P, Ponguta L, Reyes C, Karnati R. A Systematic Review of Parenting Programmes for Young Children in Low-and Middle-income Countries. United Nations Children's Fund. 2015. URL: https://www.researchgate.net/publication/316268676_A_Systematic_Review_of_Parenting_Programmes_for_Young_Children_in_Low-_and_Middle-Income_Countries [accessed 2021-07-08]
35. Ayers S, Crawley R, Webb R, Button S, Thornton A, HABiT collaborative group. What are women stressed about after birth? Birth 2019 Dec 14;46(4):678-685. [doi: 10.1111/birt.12455] [Medline: 31612558]
36. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Edited by R Kukafka; submitted 17.12.20; peer-reviewed by A Newton-Levinson, R Poluru; comments to author 01.02.21; revised version received 10.03.21; accepted 24.05.21; published 15.07.21.

Please cite as:
Chivers BR, Garad RM, Moran LJ, Lim S, Harrison CL
Support Seeking in the Postpartum Period: Content Analysis of Posts in Web-Based Parenting Discussion Groups
J Med Internet Res 2021;23(7):e26600
URL: https://www.jmir.org/2021/7/e26600
doi:10.2196/26600
PMID:34264198

©Bonnie R Chivers, Rhonda M Garad, Lisa J Moran, Siew Lim, Cheryce L Harrison. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing a Time-Adaptive Prediction Model for Out-of-Hospital Cardiac Arrest: Nationwide Cohort Study in Korea

Ji Woong Kim[1], MS; Juhyung Ha[2], MSc; Taerim Kim[3], MD; Hee Yoon[3], MD; Sung Yeon Hwang[3], MD; Ik Joon Jo[3], MD; Tae Gun Shin[3], MD; Min Seob Sim[3], MD; Kyunga Kim[1,4], PhD; Won Chul Cha[1,3,5], MD

[1]Department of Digital Health, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[2]Department of Computer Science, Indiana University, Bloomington, IN, United States
[3]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[4]Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[5]Health Information and Strategy Center, Samsung Medical Center, Seoul, Republic of Korea

Corresponding Author:
Won Chul Cha, MD
Department of Digital Health
Samsung Advanced Institute for Health Science & Technology
Sungkyunkwan University
115, Irwon-ro, Gangnam-gu, Seoul, Republic of Korea
Seoul
Republic of Korea
Phone: 82 2 3410 2053
Fax: 82 2 3410 0012
Email: docchaster@gmail.com

Abstract

Background: Out-of-hospital cardiac arrest (OHCA) is a serious public health issue, and predicting the prognosis of OHCA patients can assist clinicians in making decisions about the treatment of patients, use of hospital resources, or termination of resuscitation.

Objective: This study aimed to develop a time-adaptive conditional prediction model (TACOM) to predict clinical outcomes every minute.

Methods: We performed a retrospective observational study using data from the Korea OHCA Registry in South Korea. In this study, we excluded patients with trauma, those who experienced return of spontaneous circulation before arriving in the emergency department (ED), and those who did not receive cardiopulmonary resuscitation (CPR) in the ED. We selected patients who received CPR in the ED. To develop the time-adaptive prediction model, we organized the training data set as ongoing CPR patients by the minute. A total of 49,669 patients were divided into 39,602 subjects for training and 10,067 subjects for validation. We compared random forest, LightGBM, and artificial neural networks as the prediction model methods. Model performance was quantified using the prediction probability of the model, area under the receiver operating characteristic curve (AUROC), and area under the precision recall curve.

Results: Among the three algorithms, LightGBM showed the best performance. From 0 to 30 min, the AUROC of the TACOM for predicting good neurological outcomes ranged from 0.910 (95% CI 0.910-0.911) to 0.869 (95% CI 0.865-0.871), whereas that for survival to hospital discharge ranged from 0.800 (95% CI 0.797-0.800) to 0.734 (95% CI 0.736-0.740). The prediction probability of the TACOM showed similar flow with cohort data based on a comparison with the conventional model's prediction probability.

Conclusions: The TACOM predicted the clinical outcome of OHCA patients per minute. This model for predicting patient outcomes by the minute can assist clinicians in making rational decisions for OHCA patients.

(J Med Internet Res 2021;23(7):e28361) doi:10.2196/28361

KEYWORDS

out-of-hospital cardiac arrest; Republic of Korea; machine learning; artificial intelligence; prognosis; cardiology; prediction model

XSL•FO
RenderX

Introduction

Out-of-hospital cardiac arrest (OHCA) is a common but serious public health issue [1]. Globally, 55 in 100,000 people experience OHCA every year [2]. Despite vigorous and continuous efforts, the survival rate of OHCA patients is poor in many countries. Thus, OHCA is a critical medical problem with a poor prognosis [3,4].

Predicting the prognosis of OHCA patients can assist clinicians in deciding whether to provide treatment or use hospital resources [5,6]. Appropriate decisions allow better utilization of hospital resources, reduce medical expenses considerably, and increase the availability of care for other patients [7,8]. Furthermore, determining the termination of resuscitation for OHCA patients is an important issue because of the limited resources of hospitals. Prediction-based decision-making for using mechanical circulatory support devices or for executing early coronary angiography could improve the clinical outcomes of patients [9-11].

Some studies have proposed prediction models to predict clinical outcomes of OHCA patients using Utstein data [12,13]. These studies have shown that basic Utstein variables are associated with patient clinical outcomes. However, these studies determined patient outcomes after hospitalization and not in real time. Other studies predicting the prognosis of OHCA patients also predicted the long-term outcomes of patients, such as 1-year survival and posthospital outcomes, using various laboratory results [14,15]. Our study differs from previous studies in that the decision time and the prediction available time in our study are different.

In this study, we developed a machine learning–based time-adaptive conditional prediction model (TACOM) using the concept of a time-adaptive cohort to predict outcomes among OHCA patients every minute. We compared the random forest, LightGBM, and artificial neural network algorithms to develop a precise prediction model. The time-adaptive cohort was derived from the concept of censoring. We used the TACOM, based on the concept of the time-adaptive cohort, to compare the prediction probability with the conventional model to demonstrate the possibility of predicting patients' clinical outcomes every minute during cardiopulmonary resuscitation (CPR). To the best of our knowledge, this study is the first to predict the prognosis of OHCA patients by the minute.

Methods

Data Sets

We performed a nationwide retrospective observational cohort study using data from the Korea OHCA Registry (KOHCAR). The KOHCAR was constructed by the Korea Centers for Disease Control and Prevention (CDC) in collaboration with the Central Fire Services (CFS). We integrated the emergency medical services (EMS) run sheet, EMS CPR registry, and dispatch CPR registry into the EMS-assessed cardiac arrest database of the CFS [16,17]. Data collection was based on the Utstein style and Resuscitation Outcome Consortium Project customized for local conditions. To assess the quality of data, we held monthly meetings with field investigators and the CDC data quality control team [18]. Trained managers visited the hospitals to review the medical records and complete the database. Additionally, they contacted the patients to verify information about the outcomes [19-21]. The Korea CDC approved the use of all data.

In this study, the KOHCAR data set from January 1, 2013, to December 31, 2017, was used for the training set, and the data set from January 1, 2018, to December 31, 2018, was used for the test set. Patients who experienced return of spontaneous circulation (ROSC) before arriving in the emergency department (ED), those who did not receive CPR in the ED, and those with missing information were excluded. The institutional review board of the Samsung Medical Center approved this study.

Predictor Variables and Endpoints

Predictor variables included patient demographics, occurrence-related information, and hospital treatment information available at the time of the patient's arrival in the ED. The demographics included age and sex. Occurrence-related information included place (public or private), OHCA etiology, witness of the event, bystander CPR, prehospital CPR, patient's act at the time of OHCA, prehospital electrocardiography (ECG) rhythm, prehospital defibrillation, and history of hypertension, diabetes, heart disease, renal disease, respiratory disease, stroke, and dyslipidemia. Hospital treatment information included EMS-to-ED time, initial ECG rhythm at the ED, defibrillation, and the place of the first defibrillation.

The outcomes of this model were patient survival to hospital discharge and a good neurological outcome. For patient survival to hospital discharge, we considered patients whose ED treatment resulted in being discharged or whose hospitalization resulted in being discharged, voluntarily discharged, or transferred. A good neurological outcome was defined as a cerebral performance category of 1 or 2.

Data Processing

We used both numerical and categorical variables. The patient age and EMS-to-ED time were numerical variables. A scaling method was used for numerical variables to increase the ability of the prediction model. For the age values, we used a standard scaler that could normalize each feature by removing its mean and scaling its variance to 1. For EMS-to-ED time values, we used a robust scaler that utilized quantile information to scale each feature through the application of an inverse cumulative distribution function. For categorical variables, we used one-hot encoding to remove integer-encoded variables and then added new binary variables for each unique integer value.

Model Development

To develop a real-time outcome prediction model, we included the per minute data of patients with ongoing CPR. When predicting the clinical outcome of patients in real time, we deemed it unreasonable to predict the outcome of patients whose conditions had been determined. In other words, it is reasonable to predict the outcome of patients whose condition has not been previously determined [16]. We trained the TACOM with every single data set by minute. Data set D(t) was defined as follows:

XSL•FO
RenderX

$$D\ (T_2|T_1>t)\ \textbf{(1)}$$

where t is a minute from 0 to 60, T_1>t indicates patients whose CPR duration is longer than time t, and T_2 indicates patients who had a clinical outcome (survival to hospital discharge or good neurological recovery). The TACOM system included 61 models trained with different data sets over time. CPR duration was not used as a machine learning feature in the model. The model's training data set cohort varied for each minute according to the duration of CPR.

Additionally, we developed three models, namely, random forest, LightGBM, and artificial neural networks, to select the best performance model. Both the deep learning algorithm and machine learning algorithm performed well in the health care domain [22,23]. We compared the area under the receiver operating characteristic curve (AUROC) for model performance and chose LightGBM as our final model. We have provided the AUROC of the other models in Multimedia Appendix 1. LightGBM, an open-source algorithm by Microsoft, is an advanced model of the ensemble algorithm for speeding up the training process and reducing memory consumption. In general, the ensemble model has shown remarkable performance for the classification of structured data. It generates several classifiers and combines predictions to derive a final prediction. Ensembles have the following two main types: bagging and boosting. In the bagging algorithm, each training set is constructed by forming a bootstrap replicate of the original training set. In the boosting algorithm, the model maintains a set of weights over the original training set and adjusts these weights after each classifier is learned by the base algorithm [24].

We utilized a widely used parameter optimization algorithm, the grid search, to determine the best combination of the three hyperparameters in LightGBM. Referring to the technical documents provided by Microsoft, first, we found the value of "max_depth," which specifies the depth limit of the tree. Through grid search, we selected the value that had the highest AUROC among the values from 1 to 10. After that, we found the value of "num_leaves" that controls the complexity of the tree model. Theoretically, since the number of leaves should be smaller than $2^{\text{"max_depth"}}$, one of the values from 100 to 900 was selected through the grid search. Finally, we found "min_data_in_leaf," an essential parameter to prevent over-fitting. Its value ranges from 100 to 1000; we found an appropriate value through the grid search. The AUROC of each value is shown in Multimedia Appendix 2.

Statistical Analysis

For the descriptive statistics, means and SDs were used for continuous variables, and frequencies and percentages were used for categorical variables. The *t* test and chi-squared test were performed to determine the mean differences between the derivation and test sets. All tests were two-tailed with the statistical significance level set at P<.05. Additionally, the standardized mean difference (SMD) was used to measure the effect size of the two groups.

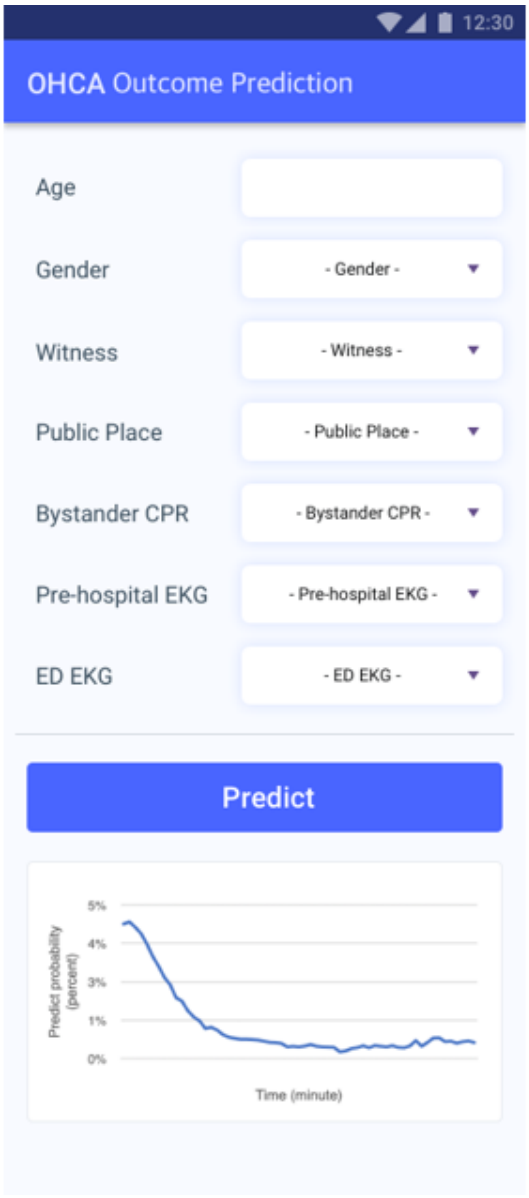
We used various metrics, including prediction probability, AUROC, and area under the precision-recall curve (AUPRC), to measure the metric of our prediction model, TACOM. Prediction probability was used to determine which model reflected reality. AUROC and AUPRC scores were used to measure the performance of the binary outcomes. We evaluated 95% CIs using bootstrapping with 1000 sampling iterations with replacement.

Implementation

Furthermore, we developed a simple user interface (Figure 1) for showing prediction probability using Android. By simply entering the input values, a patient's outcomes can be predicted and visualized as a graph. We designed an application prototype. This application could provide information to the medical staff. The implemented software for model development included the Python programming language (version 3.8.5), Tensorflow framework (version 2.3.1), and scikit-learn (version 0.23.2). Using the Tensorflow framework, the predictive model could be extended to a mobile or web application.

XSL•FO
RenderX

Figure 1. Simple user interface of the out-of-hospital cardiac arrest outcome prediction model.



Code Availability

We published our prediction model on GitHub [25]. The codes that support the findings of this study are available on GitHub.

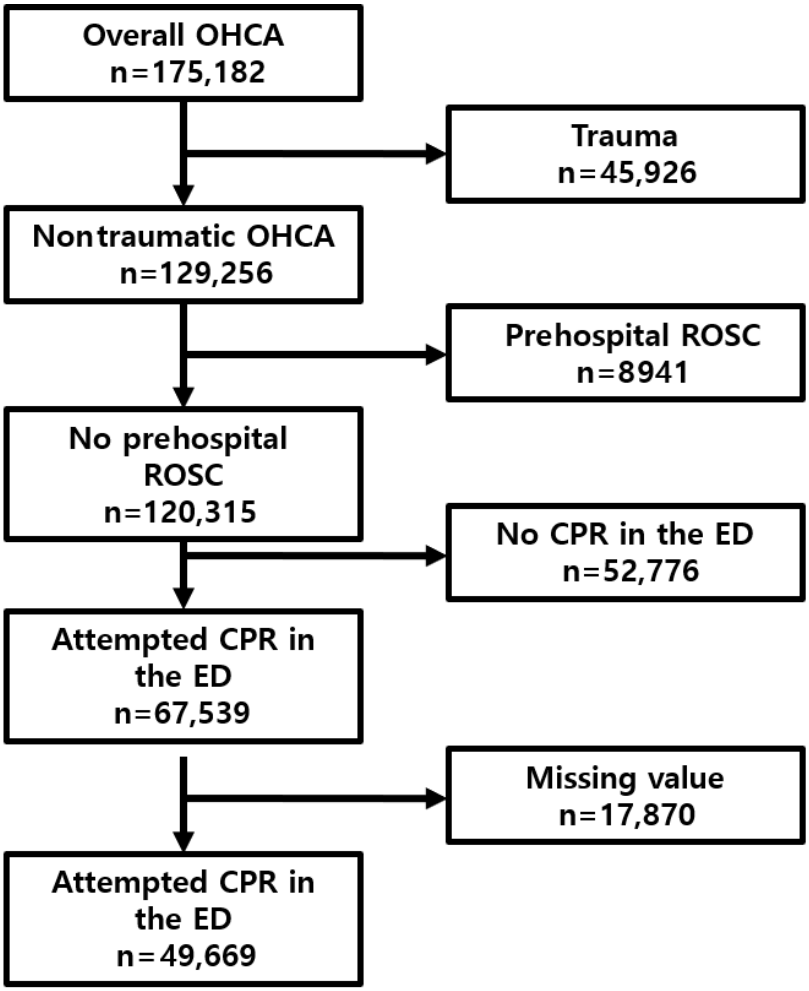
Results

Patient Selection

Patients' records from the KOHCAR were used as derivation and validation data sets. The KOHCAR held records for 175,182 patients from 2013 to 2018. After excluding trauma patients, patients who experienced ROSC in the prehospital stage, patients who did not receive CPR in the ED, and patients with missing values, we included 49,669 records in our study. We split the data set into 39,602 records from 2013 to 2017 to be used as the derivation set, and 10,067 records from 2018 to be used as the validation set. Figure 2 presents the diagram of patient selection from the KOHCAR.

XSL•FO
RenderX

Figure 2. Subject selection process of OHCA patients. CPR: cardiopulmonary resuscitation; ED: emergency department; OHCA: out-of-hospital cardiac arrest; ROSC: return of spontaneous circulation.



Patient Characteristics

Table 1 shows the baseline characteristics of the derivation and validation sets from the KOHCAR. Overall, the patients' basic characteristics showed that the majority were men (32,049/49,669, 64.5%), and the average age was 67.0 years, with a quartile range of 57.0 to 79.0 years. The shockable rhythm rates in the ED and EMS were 4.4% (2170/49,669) and 10.2% (5,067/49,669), respectively.

XSL•FO
RenderX

Table 1. Basic characteristics of the study participants.

Variable	All (n=49,669)	Derivation data (n=39,602)	Validation data (n=10,067)	*P* value	SMD[a]
Age, mean (SD)	67.0 (18.8)	66.5 (19.3)	68.8 (16.6)	<.001	0.125
Female sex, n (%)	17,620 (35.5%)	13,932 (35.2%)	3688 (36.6%)	.007	0.030
Public place, n (%)	8030 (16.2%)	6517 (16.5%)	1513 (15.0%)	<.001	0.130
Witnessed, n (%)	30,314 (61.0%)	24,126 (60.9%)	3552 (35.3%)	<.001	0.190
Bystander CPR[b], n (%)	9966 (20.1%)	7417 (18.7%)	2549 (25.3%)	<.001	0.333
Cause, n (%)				<.001	0.114
Cardiogenic disease	45,792 (92.2%)	36,495 (92.2%)	9297 (92.4%)		
Respiratory disease	381 (0.8%)	303 (0.8%)	78 (0.8%)		
Nontraumatic bleeding	775 (1.6%)	558 (1.4%)	217 (2.2%)		
Terminal cancer	459 (0.9%)	430 (1.1%)	29 (0.3%)		
Sudden infant death syndrome	197 (0.4%)	147 (0.4%)	50 (0.5%)		
Others	2065 (4.2%)	1,669 (4.2%)	396 (3.9%)		
Initial ECG[c] rhythm of EMS[d], n (%)				<.001	N/A[e]
VF[f]	4732 (9.5%)	3603 (9.1%)	1129 (11.2%)		
Pulseless VT[g]	335 (0.7%)	258 (0.7%)	77 (0.8%)		
Asystole	15,820 (31.9%)	11,593 (29.3%)	4227 (42.0%)		
PEA[h]	5406 (10.9%)	3637 (9.2)	1769 (17.6%)		
Others	23,376 (47.0%)	20,511 (51.7%)	2865 (28.4%)		
Initial ECG rhythm of ED[i], n (%)				<.001	N/A
VF	1913 (3.9%)	1576 (4.0%)	337 (3.3%)		
Pulseless VT	257 (0.5%)	218 (0.6%)	39 (0.4%)		
Asystole	29,433 (59.3%)	23,532 (59.4%)	5901 (58.6%)		
PEA	5615 (11.3%)	4031 (10.2%)	1584 (15.7%)		
Others	12,451 (25.0%)	10,245 (25.8%)	2206 (22.0%)		
Anamnesis hypertension, n (%)	17,709 (35.7%)	13,886 (35.1%)	3823 (38.0%)	<.001	0.184
Anamnesis diabetes, n (%)	11,787 (23.7%)	9188 (23.2%)	2599 (25.8%)	<.001	0.209
Anamnesis heart disease, n (%)	8720 (17.6%)	6774 (17.1%)	1946 (19.3%)	<.001	0.059
Anamnesis renal disease, n (%)	3192 (6.4%)	2489 (6.3%)	703 (7.0%)	.02	0.031
Anamnesis respiratory disease, n (%)	3373 (6.8%)	2613 (6.6%)	760 (7.5%)	.003	0.037
Anamnesis stroke, n (%)	4261 (8.6%)	3333 (8.4%)	928 (9.2%)	.01	0.033
Anamnesis dyslipidemia, n (%)	1201 (2.4%)	863 (2.2%)	338 (3.4%)	<.001	0.073
EMS arrival (min), mean (SD)	45.9 (75.5)	44.7 (73.0)	50.6 (84.5)	<.001	0.074

[a]SMD: standardized mean difference.

[b]CPR: cardiopulmonary resuscitation.

[c]ECG: electrocardiography.

[d]EMS: emergency medical services.

[e]N/A: not applicable.

[f]VF: ventricular fibrillation.

[g]VT: ventricular tachycardia.

[h]PEA: pulseless electrical activity.

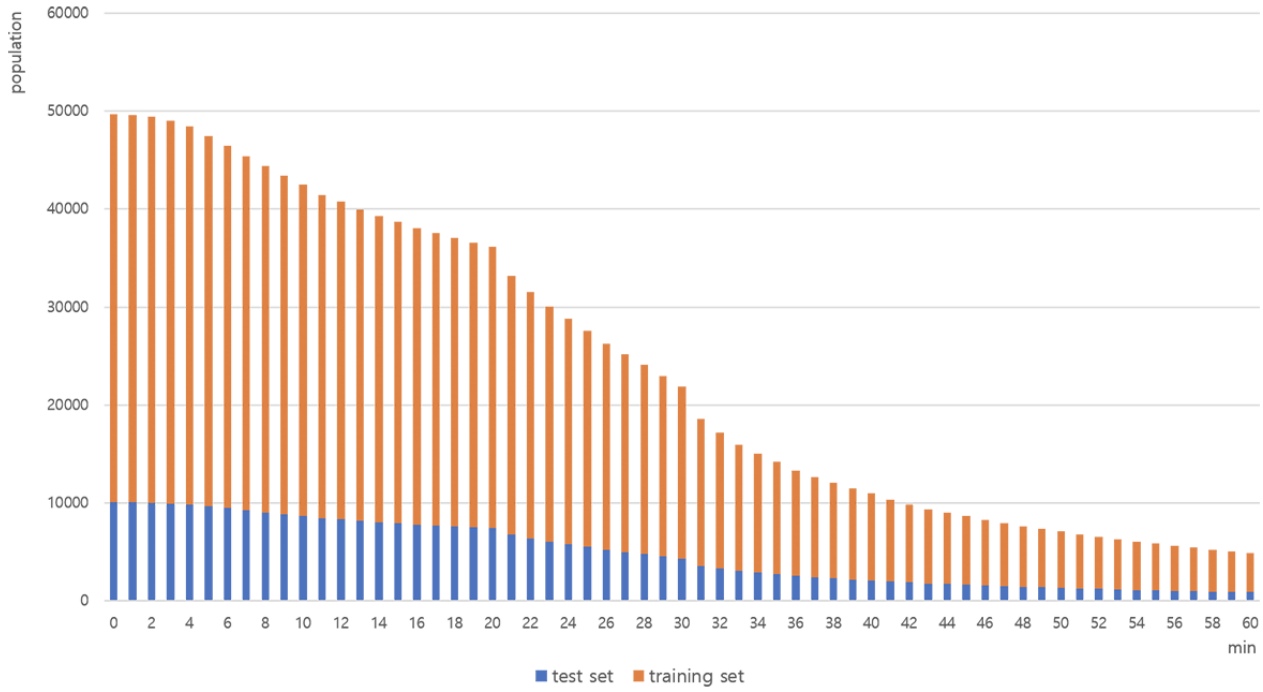
[i]ED: emergency department.

XSL•FO
RenderX

Study Cohort

As shown in Figure 3, we divided our data set by CPR duration from 0 to 60 minutes. As time progressed, the models were trained using patients who were receiving CPR. In other words, patients who did not receive CPR were excluded. We did not use CPR duration as a feature but as a criterion that divided the data by minutes to apply the concept of time. The first model included 49,669 patients. The 30-minute model was trained using the data of 21,841 patients, indicating that the status of 27,828 patients in the ED had already been determined after 30 minutes.

Figure 3. Population of every data set included in each minute from 0 to 60.



Model Performance

Figure 4 shows the prediction probability of the outcome (survival to hospital discharge) for the TACOM and that for a prediction model that uses CPR duration as a machine learning feature. Considering that the survival to hospital discharge rate of OHCA patients is about 5%, the TACOM reflected reality, whereas the prediction model that used CPR duration as a training feature was overly optimistic.

Figure 5 shows the receiver operating characteristic and precision-recall curves for the test set at 2 minutes, trained with data on patients whose outcomes had not been determined before 2 minutes. The AUROC and AUPRC of all the minutes from 0 to 30 are shown in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 4. Prediction probability of the TACOM and conventional model for out-of-hospital cardiac arrest patients' survival to hospital discharge. TACOM: time-adaptive conditional prediction model.

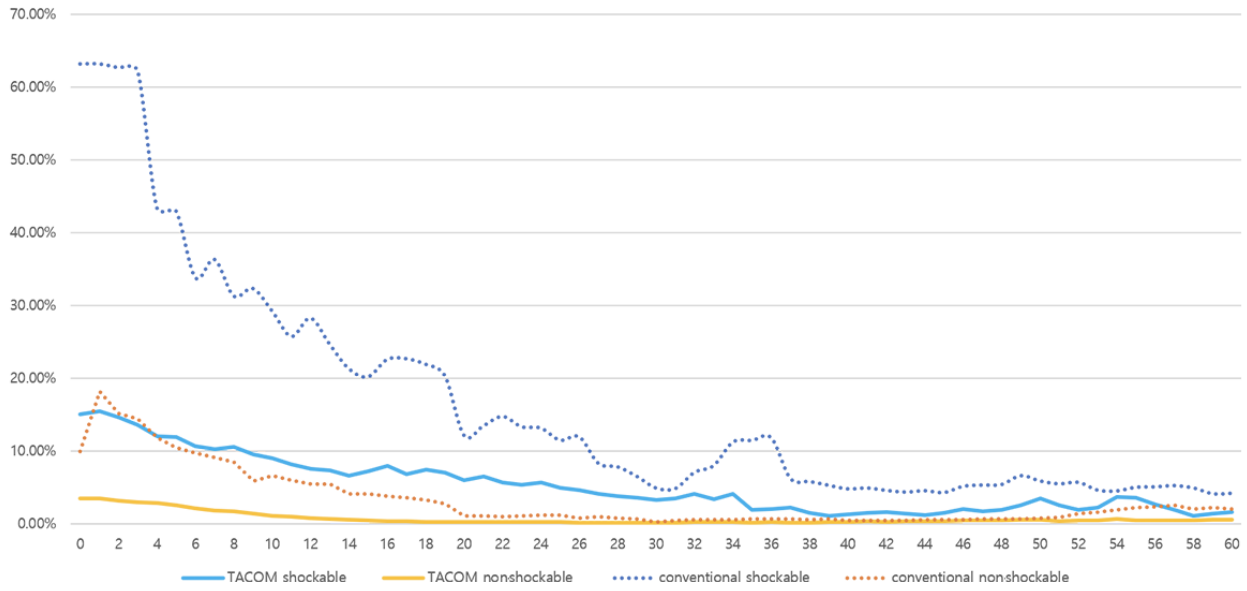
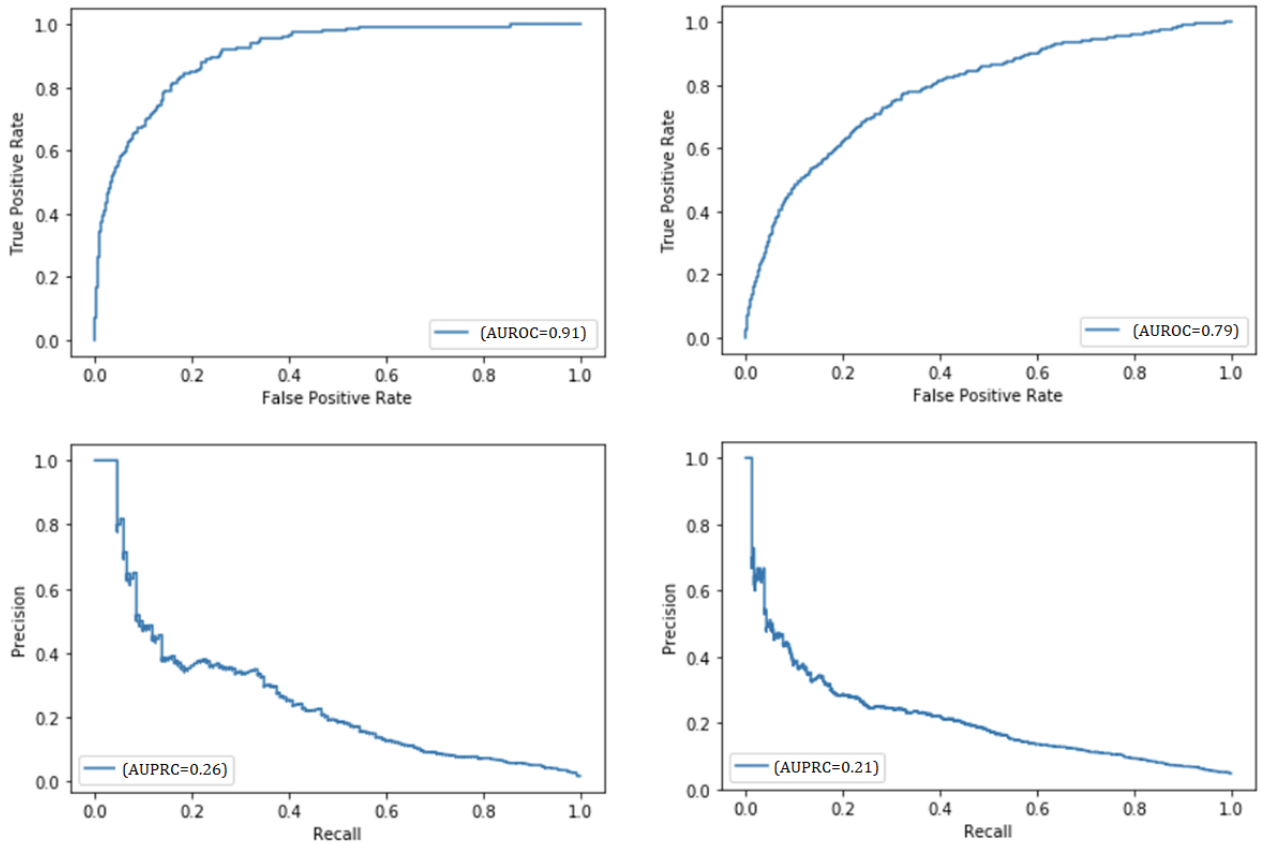


Figure 5. Area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) of the time-adaptive conditional model for out-of-hospital cardiac arrest (OHCA) patients' survival to hospital discharge at 2 minutes (right) and OHCA patients' good neurological outcome at 2 minutes (left).



XSL•FO
RenderX

Discussion

Principal Findings

We developed a time-adaptive conditional model for predicting the clinical outcomes of OHCA patients using machine learning data from a large nationwide cohort registry. Our study demonstrated the possibility of real-time prediction among OHCA patients using a time-adaptive conditional model. The most important finding of this study was that the time-adaptive conditional model, in which CPR duration was not used as a feature but as a criterion to define the time-adaptive cohort, reflected real-world outcomes. The training data were divided by time according to the duration of CPR and trained separately for each time. Thus, the model was verified to be suitable for the real-time prediction of clinical outcomes.

The TACOM predicted the probability of survival of specific patients, updated every minute in a personalized manner. Hence, it could have practical application in the field. Currently, clinical decision-making is based on the personal experience of the medical staff or institutional guidelines; thus, decisions do not necessarily reflect the actual condition of each patient [17]. We used a machine learning–based model to calculate a patient's clinical outcome in real time using the patient's initial state. The clinical outcome prediction probability of the TACOM showed a similar flow to the actual clinical outcome rate of the cohort data.

We designed the TACOM to be different from the conventional model; the TACOM only used the initial information that could be obtained before CPR as a machine learning feature. Additionally, data, such as time arrival to EMS, time to EMS arrival at the ER, and lab results, were excluded. However, unlike with other models, the TACOM reflects the real environment using a time-adaptive cohort. We made the time-adaptive cohort from one big registry by censoring the data. Censoring is used when time-to-event information is not available such as in clinical trials or survival analysis in cancer treatment. We did not train all patients at once; however, we created the discriminative models by censoring the patients whose status was determined by the minute.

We aimed to apply and test the TACOM in the real world to make it practically useful and effective in the future. In this study, validation was performed using a subset of the data set. Prospective data collection and verification would be required and potentially achieved by developing an application that applies the TACOM. Furthermore, usability and utility evaluations for UX design are needed to identify whether the application is convenient and useful.

Limitations

This study had some limitations. First, the cohort was organized in the Utstein style. The data were obtained from a nationwide cohort, and thus, some detailed characteristics reflecting the quality of CPR and patient responses were unavailable. Second, our model did not include long-term outcomes, such as 1-month or 1-year survival. Finally, given the absence of any significant difference between machine learning algorithms, various machine learning algorithms have not been included in this study.

Conclusions

We developed a time-adaptive conditional model to predict the clinical outcomes of OHCA patients per minute. We found a suitable algorithm for the TACOM to predict the survival to hospital discharge and neurological recovery of OHCA patients at the minute level. This study showed the potential of the time-adaptive prediction model for resuscitation, which can be useful to medical staff for making appropriate and rational decisions.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI19C0275010021).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Area under the receiver operating characteristic curve of the time-adaptive conditional model using the following three different methods: LightGBM, random forest, and deep learning.
[DOCX File , 19 KB - jmir_v23i7e28361_app1.docx]

Multimedia Appendix 2
Results of the grid search according to the three hyperparameters in LightGBM.
[DOCX File , 17 KB - jmir_v23i7e28361_app2.docx]

Multimedia Appendix 3
The area under the receiver operating characteristic curve and area under the precision-recall curve of the time-adaptive conditional model from 0 to 30 minutes.

XSL•FO
RenderX

[DOCX File , 16 KB - jmir_v23i7e28361_app3.docx]

References

1. McCarthy JJ, Carr B, Sasson C, Bobrow BJ, Callaway CW, Neumar RW, et al. Out-of-Hospital Cardiac Arrest Resuscitation Systems of Care: A Scientific Statement From the American Heart Association. Circulation 2018 May 22;137(21):645-660. [doi: 10.1161/cir.0000000000000557]
2. Yan S, Gan Y, Jiang N, Wang R, Chen Y, Luo Z, et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. Crit Care 2020 Feb 22;24(1):61-13 [FREE Full text] [doi: 10.1186/s13054-020-2773-2] [Medline: 32087741]
3. Hreinsson JP, Thorvaldsson AP, Magnusson V, Fridriksson BT, Libungan BG, Karason S. Identifying out-of-hospital cardiac arrest patients with no chance of survival: An independent validation of prediction rules. Resuscitation 2020 Jan 01;146:19-25. [doi: 10.1016/j.resuscitation.2019.11.001] [Medline: 31711916]
4. Martinell L, Nielsen N, Herlitz J, Karlsson T, Horn J, Wise MP, et al. Early predictors of poor outcome after out-of-hospital cardiac arrest. Crit Care 2017 Apr 13;21(1):96 [FREE Full text] [doi: 10.1186/s13054-017-1677-2] [Medline: 28410590]
5. Carrick RT, Park JG, McGinnes HL, Lundquist C, Brown KD, Janes WA, et al. Clinical Predictive Models of Sudden Cardiac Arrest: A Survey of the Current Science and Analysis of Model Performances. JAHA 2020 Aug 18;9(16):1-16. [doi: 10.1161/jaha.119.017625]
6. Kim HS, Park KN, Kim SH, Lee BK, Oh SH, Jeung KW, et al. Prognostic value of OHCA, C-GRApH and CAHP scores with initial neurologic examinations to predict neurologic outcomes in cardiac arrest patients treated with targeted temperature management. PLoS One 2020 Apr 24;15(4):e0232227 [FREE Full text] [doi: 10.1371/journal.pone.0232227] [Medline: 32330180]
7. Skrifvars M, Vayrynen T, Kuisma M, Castren M, Parr M, Silfverstople J, et al. Comparison of Helsinki and European Resuscitation Council "do not attempt to resuscitate" guidelines, and a termination of resuscitation clinical prediction rule for out-of-hospital cardiac arrest patients found in asystole or pulseless electrical activity. Resuscitation 2010 Jun;81(6):679-684. [doi: 10.1016/j.resuscitation.2010.01.033] [Medline: 20381229]
8. Goto Y, Maeda T, Goto Y. Termination-of-resuscitation rule for emergency department physicians treating out-of-hospital cardiac arrest patients: an observational cohort study. Crit Care 2013 Oct 13;17(5):R235 [FREE Full text] [doi: 10.1186/cc13058] [Medline: 24119782]
9. Atkinson TM, Ohman EM, O'Neill WW, Rab T, Cigarroa JE, Interventional Scientific Council of the American College of Cardiology. A Practical Approach to Mechanical Circulatory Support in Patients Undergoing Percutaneous Coronary Intervention: An Interventional Perspective. JACC Cardiovasc Interv 2016 May 09;9(9):871-883 [FREE Full text] [doi: 10.1016/j.jcin.2016.02.046] [Medline: 27151604]
10. Asleh R, Resar JR. Utilization of Percutaneous Mechanical Circulatory Support Devices in Cardiogenic Shock Complicating Acute Myocardial Infarction and High-Risk Percutaneous Coronary Interventions. J Clin Med 2019 Aug 13;8(8):1209 [FREE Full text] [doi: 10.3390/jcm8081209] [Medline: 31412669]
11. Sauneuf B, Dupeyrat J, Souloy X, Leclerc M, Courteille B, Canoville B, et al. The CAHP (cardiac arrest hospital prognosis) score: A tool for risk stratification after out-of-hospital cardiac arrest in elderly patients. Resuscitation 2020 Mar 01;148:200-206. [doi: 10.1016/j.resuscitation.2020.01.011] [Medline: 31987887]
12. Rea TD, Cook AJ, Stiell IG, Powell J, Bigham B, Callaway CW, Resuscitation Outcomes Consortium Investigators. Predicting survival after out-of-hospital cardiac arrest: role of the Utstein data elements. Ann Emerg Med 2010 Mar;55(3):249-257. [doi: 10.1016/j.annemergmed.2009.09.018] [Medline: 19944488]
13. Baldi E, Caputo ML, Savastano S, Burkart R, Klersy C, Benvenuti C, et al. An Utstein-based model score to predict survival to hospital admission: The UB-ROSC score. Int J Cardiol 2020 Jun 01;308:84-89 [FREE Full text] [doi: 10.1016/j.ijcard.2020.01.032] [Medline: 31980268]
14. Seki T, Tamura T, Suzuki M, SOS-KANTO 2012 Study Group. Outcome prediction of out-of-hospital cardiac arrest with presumed cardiac aetiology using an advanced machine learning technique. Resuscitation 2019 Aug;141:128-135. [doi: 10.1016/j.resuscitation.2019.06.006] [Medline: 31220514]
15. Pareek N, Kordis P, Beckley-Hoelscher N, Pimenta D, Kocjancic S, Jazbec A, et al. A practical risk score for early prediction of neurological outcome after out-of-hospital cardiac arrest: MIRACLE2. Eur Heart J 2020 Dec 14;41(47):4508-4517. [doi: 10.1093/eurheartj/ehaa570] [Medline: 32731260]
16. Hong KJ, Shin SD, Song KJ, Cha WC, Cho JS. Association between ED crowding and delay in resuscitation effort. Am J Emerg Med 2013 Mar;31(3):509-515. [doi: 10.1016/j.ajem.2012.09.029] [Medline: 23159432]
17. Park CH, Ahn KO, Shin SD, Park JH, Lee SY. Association between health insurance status and transfer of patients with return of spontaneous circulation after out-of-hospital cardiac arrest. Resuscitation 2020 Apr;149:143-149. [doi: 10.1016/j.resuscitation.2020.02.018] [Medline: 32114072]
18. Kim YT, Shin SD, Hong SO, Ahn KO, Ro YS, Song KJ, et al. Effect of national implementation of utstein recommendation from the global resuscitation alliance on ten steps to improve outcomes from Out-of-Hospital cardiac arrest: a ten-year observational study in Korea. BMJ Open 2017 Aug 21;7(8):e016925 [FREE Full text] [doi: 10.1136/bmjopen-2017-016925] [Medline: 28827263]

XSL•FO
RenderX

19. Ro YS, Shin SD, Song KJ, Lee EJ, Kim JY, Ahn KO, et al. A trend in epidemiology and outcomes of out-of-hospital cardiac arrest by urbanization level: a nationwide observational study from 2006 to 2010 in South Korea. Resuscitation 2013 May;84(5):547-557. [doi: 10.1016/j.resuscitation.2012.12.020] [Medline: 23313428]
20. Yoo YJ, Kim GW, Lee CA, Park YJ, Lee KM, Cho JS, et al. Characteristics and outcomes of public bath-related out-of-hospital cardiac arrests in South Korea. Clin Exp Emerg Med 2020 Sep 30;7(3):225-233 [FREE Full text] [doi: 10.15441/ceem.19.071] [Medline: 33028067]
21. Policy Information: Health Survey. Korea Disease Control Prevention Agency. URL: http://www.kdca.go.kr/contents.es?mid=a20303010401 [accessed 2020-12-30]
22. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
23. Singh A, Thakur N, Sharma A. A review of supervised machine learning algorithms. 2016 Presented at: 3rd International Conference on Computing for Sustainable Global Development (INDIACom); March 16-18, 2016; New Delhi, India URL: https://ieeexplore.ieee.org/abstract/document/7724478/
24. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 2017; Long Beach, CA p. 3149-3157.
25. OHCA. GitHub. URL: https://github.com/jiwoongme/OHCA [accessed 2021-06-27]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
CDC: Centers for Disease Control and Prevention
CFS: Central Fire Services
CPR: cardiopulmonary resuscitation
ECG: electrocardiography
ED: emergency department
EMS: emergency medical services
KOHCAR: Korea Out-of-Hospital Cardiac Arrest Registry
OHCA: out-of-hospital cardiac arrest
ROSC: return of spontaneous circulation
TACOM: time-adaptive conditional prediction model

Edited by G Eysenbach; submitted 02.03.21; peer-reviewed by JA Benítez-Andrades, MT García-Ordás; comments to author 19.03.21; revised version received 12.05.21; accepted 31.05.21; published 05.07.21.

Please cite as:
Kim JW, Ha J, Kim T, Yoon H, Hwang SY, Jo IJ, Shin TG, Sim MS, Kim K, Cha WC
Developing a Time-Adaptive Prediction Model for Out-of-Hospital Cardiac Arrest: Nationwide Cohort Study in Korea
J Med Internet Res 2021;23(7):e28361
URL: https://www.jmir.org/2021/7/e28361/
doi:10.2196/28361
PMID:36260382

©Ji Woong Kim, Juhyung Ha, Taerim Kim, Hee Yoon, Sung Yeon Hwang, Ik Joon Jo, Tae Gun Shin, Min Seob Sim, Kyunga Kim, Won Chul Cha. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effective Treatment Recommendations for Type 2 Diabetes Management Using Reinforcement Learning: Treatment Recommendation Model Development and Validation

Xingzhi Sun[1*], PhD; Yong Mong Bee[2,3*], MBBS, MRCP; Shao Wei Lam[4,5], MEng, PhD; Zhuo Liu[1], MSc; Wei Zhao[1], PhD; Sing Yi Chia[6], BSc; Hanis Abdul Kadir[5,6], BSc; Jun Tian Wu[4], MPH; Boon Yew Ang[4], MTech; Nan Liu[4,5], PhD; Zuo Lei[1], MSc; Zhuoyang Xu[1], MSc; Tingting Zhao[1], MSc; Gang Hu[1], PhD; Guotong Xie[1,7,8], PhD

[1]Ping An Healthcare Technology, Beijing, China

[2]Department of Endocrinology, Singapore General Hospital, Singapore, Singapore

[3]SingHealth Duke-NUS Diabetes Centre, Singapore Health Services, Singapore, Singapore

[4]Health Services Research Centre, Singapore Health Services, Singapore, Singapore

[5]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

[6]Health Services Research Unit, Singapore General Hospital, Singapore, Singapore

[7]Ping An Healthcare and Technology Co, Ltd, Shanghai, China

[8]Ping An International Smart City Technology Co, Ltd, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Guotong Xie, PhD
Ping An Healthcare Technology
Floor 9, Building B
Ping An International Finance Center
Beijing, 100026
China
Phone: 86 15201521697
Email: xieguotong@pingan.com.cn

Abstract

Background: Type 2 diabetes mellitus (T2DM) and its related complications represent a growing economic burden for many countries and health systems. Diabetes complications can be prevented through better disease control, but there is a large gap between the recommended treatment and the treatment that patients actually receive. The treatment of T2DM can be challenging because of different comprehensive therapeutic targets and individual variability of the patients, leading to the need for precise, personalized treatment.

Objective: The aim of this study was to develop treatment recommendation models for T2DM based on deep reinforcement learning. A retrospective analysis was then performed to evaluate the reliability and effectiveness of the models.

Methods: The data used in our study were collected from the Singapore Health Services Diabetes Registry, encompassing 189,520 patients with T2DM, including 6,407,958 outpatient visits from 2013 to 2018. The treatment recommendation model was built based on 80% of the dataset and its effectiveness was evaluated with the remaining 20% of data. Three treatment recommendation models were developed for antiglycemic, antihypertensive, and lipid-lowering treatments by combining a knowledge-driven model and a data-driven model. The knowledge-driven model, based on clinical guidelines and expert experiences, was first applied to select the candidate medications. The data-driven model, based on deep reinforcement learning, was used to rank the candidates according to the expected clinical outcomes. To evaluate the models, short-term outcomes were compared between the model-concordant treatments and the model-nonconcordant treatments with confounder adjustment by stratification, propensity score weighting, and multivariate regression. For long-term outcomes, model-concordant rates were included as independent variables to evaluate if the combined antiglycemic, antihypertensive, and lipid-lowering treatments had a positive impact on reduction of long-term complication occurrence or death at the patient level via multivariate logistic regression.

Results: The test data consisted of 36,993 patients for evaluating the effectiveness of the three treatment recommendation models. In 43.3% of patient visits, the antiglycemic medications recommended by the model were concordant with the actual

XSL•FO
RenderX

prescriptions of the physicians. The concordant rates for antihypertensive medications and lipid-lowering medications were 51.3% and 58.9%, respectively. The evaluation results also showed that model-concordant treatments were associated with better glycemic control (odds ratio [OR] 1.73, 95% CI 1.69-1.76), blood pressure control (OR 1.26, 95% CI, 1.23-1.29), and blood lipids control (OR 1.28, 95% CI 1.22-1.35). We also found that patients with more model-concordant treatments were associated with a lower risk of diabetes complications (including 3 macrovascular and 2 microvascular complications) and death, suggesting that the models have the potential of achieving better outcomes in the long term.

Conclusions: Comprehensive management by combining knowledge-driven and data-driven models has good potential to help physicians improve the clinical outcomes of patients with T2DM; achieving good control on blood glucose, blood pressure, and blood lipids; and reducing the risk of diabetes complications in the long term.

(J Med Internet Res 2021;23(7):e27858) doi:10.2196/27858

KEYWORDS

type 2 diabetes; reinforcement learning; model concordance; short-term outcome; long-term outcome

Introduction

Type 2 diabetes mellitus (T2DM) is a worldwide chronic disease characterized by higher than optimal blood glucose levels. T2DM can lead to multiple complications and increase the risk of death. According to the global report on diabetes of the World Health Organization [1], 3.7 million people died of diabetes in 2012 and the prevalence has been increasing in the past three decades. T2DM and its related complications represent a growing economic burden for many countries and health systems [2]. Diabetes complications can be prevented through better disease control; however, there is still a large gap between the recommended treatment and the treatment that patients actually receive [3].

The treatment of T2DM can be challenging because of the different therapeutic targets and individual variability of the patients, leading to the need for precise, personalized treatment [4]. In addition, patients with diabetes require a sequence of treatments due to chronicity of the condition, each of which may affect the patients' clinical outcome in the long term. The decision-making for determining a sequence of treatments can be more complex because (1) the impact of a single treatment may not be immediately reflected, and (2) if we regard all of the treatments a patient received chronically as a treatment program, the number of options for the treatment programs is extremely large and finding the best program for an individual patient is a great challenge.

With the explosive increase of electronic medical records (EMRs) and the rapid development of artificial intelligence technology, it has now become possible to teach a model that enables personalized treatment with the best expected clinical outcomes. The treatment of chronic diseases such as T2DM is a sequential decision-making process. Our goal is to develop effective treatment regimens that can dynamically adapt to the varying clinical states and maximize the long-term benefits of patients. Reinforcement learning (RL) [5] is an approach that learns the best policy toward a predefined long-term goal via trial and error to address a sequential decision-making problem. The RL approach has intrinsic advantages of tackling the treatment recommendation problem for chronic diseases. First, by considering the accumulative rewards as the optimization goal, the long-term effect of current decision-making is taken into account. Second, the design of RL leverages all samples in model development by reinforcing actions with a good reward and punishing others with a bad reward. With theoretical and technical developments in recent years, the RL approach has been successfully applied in the health care domain, including for chronic disease management [6-10], critical care [11-14], and other forms of health management [15,16].

The treatment of chronic diseases consists of a sequence of medications or procedures that are determined based on the changing clinical conditions of a patient and the effects from the previous treatment. Tseng et al [6] recently proposed an RL-based model to automate adaptive radiotherapy decision-making for patients with nonsmall cell lung cancer, where the deep Q network (DQN) was used to learn dose decisions based on real clinical data and the synthesized data created by generative adversarial networks [17]. The framework was evaluated in a dataset of 114 patients. The learned dose strategies by the DQN could achieve similar results to those decided by clinicians, yielding feasible and promising solutions for automatic treatment designs. Once a treatment recommendation model is developed, it is imperative to carefully evaluate its validity and effectiveness before wide application. In the clinical domain, a randomized controlled trial (RCT) is often performed to test the efficacy of an intervention. However, RCTs can be costly, unpractical, and infeasible in some clinical scenarios [18]. With the increase of EMR use, a retrospective study has become a reasonable alternative to evaluate models via statistical tests and other data analytics methods.

There is an emerging trend in the literature for effectiveness evaluation on the treatment of chronic diseases [19-21], such as the comparative effectiveness of more or less aggressive treatment intensification strategies in adults with T2DM [19]. In these studies, two types of treatments are compared in terms of a short-term clinical outcome such as the key indicator of the disease and a long-term outcome such as the occurrences of complications or death. When making a comparison in such observational studies, it is crucial to eliminate the influence of confounding factors. For short-term clinical outcomes, multiple logistic regression and the propensity score (PS) method are conventional approaches to adjust the confounders between treatment groups [22]. With respect to long-term outcomes, survival analysis via the Cox proportional hazard model can be applied to adjust the time-invariant or time-varying covariates for two treatment groups [23]. If the covariates change over

XSL•FO
RenderX

time and are affected by the previous treatment, the use of marginal structural models [24-26] was proposed to control the confounders. In relation to diabetes treatment, Chen et al [27] assessed the association between treatment concordance with clinical guidelines and related clinical outcomes in patients with T2DM by comparing guideline-concordant and guideline-nonconcordant cohorts. This work is closely related to treatment model evaluation as the treatment groups to be compared are defined based on a given guideline, which can be regarded as a special treatment model that has been verified and commonly accepted. Chen et al [27] considered hospital admission and severe hypoglycemic events as the clinical outcomes of interest. Logistic regressions were used to examine factors associated with the likelihood of having at least one hospital admission and Cox proportional hazard regressions were used to model time to hypoglycemic events.

In this work, we developed treatment recommendation models based on the deep RL approach and then performed a retrospective study to evaluate the reliability and effectiveness of the models. The anonymized data used in our study are derived from the Singapore Health Services (SingHealth) Diabetes Registry [28], which is built based on the EMRs from SingHealth, the largest health cluster in Singapore with 4 hospitals, 5 national centers, 8 polyclinics (primary care clinics), and 3 intermediate long-term-care community hospitals. For treatment recommendation, we successfully applied the deep RL technique in the context of personalized treatment for patients with T2DM, with careful design and formulation for this challenging problem. We built a model that can be used to recommend the medications for patients with T2DM based on their clinical information, including demographic data, vital signs, laboratory tests, disease history, and current medications. Three models were developed for antiglycemic, antihypertensive, and lipid-lowering treatments to enable the comprehensive management of patients with T2DM. We evaluated the effectiveness of our treatment recommendation models by systematically performing a retrospective study on the EMRs of patients with diabetes.

Methods

Patient Characteristics

This retrospective study was based on anonymized data of 189,520 patients with T2DM from SingHealth Diabetes Registry between January 2013 and December 2018. The study was approved by the SingHealth Centralized Institutional Review Board with a waiver of informed consent granted. The board deemed that further ethical deliberation was not required as the study involves analysis of an anonymized dataset. All methods performed in this study were in accordance with the relevant guidelines and regulations. The dataset was split into training data (80% with 152,527 patients) for treatment recommendation models, including three types of treatments (antiglycemic, antihypertensive, and lipid-lowing treatment), and test data (20% of data with 36,993 patients) for evaluating the effectiveness of the three treatment recommendation models.

The EMR data for each patient included demographic information, medical history, physical measurements, laboratory data, and physicians' prescriptions. Demographic information included age, gender, ethnicity, smoking, and others. Medical history included comorbidities, vascular complications, hospital admissions, emergency department visits, and outpatient visits. Physical measurements included systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, weight, height, and BMI. Laboratory data included glycated hemoglobin A_{1c} (HbA_{1c}), low-density lipoprotein cholesterol (LDL-c), and fasting plasma glucose. For physicians' prescriptions, we considered only antiglycemic, antihypertensive, and lipid-lowering drugs and their dosages.

As is the case for all EMR data, our dataset contains errors and missing data. The rate of missingness was generally low, with higher rates for variables under the categories of physical measurements and laboratory data. We handled the errors and missing data using the following strategy. During preprocessing, errors were treated as missing values. For missing physical measurements and laboratory data, we substituted the missing data with the value from the closest preceding data point of the same patient within a 1-year time frame. If data were still missing, we proceeded to impute the missing data using the median of the observed values for that variable for all patients without missing data.

Clinical Outcomes

Two types of clinical outcomes were analyzed: short-term and long-term outcomes. Short-term outcomes were evaluated at the patient-visit level, including blood glucose control, blood pressure control, blood lipids control, and hypoglycemia-related admissions. For the long-term outcomes, we evaluated the occurrences of 5 diabetes complications and death in up to 6 years at the patient level, including myocardial infarction, heart failure, stroke (including ischemic and hemorrhagic strokes), diabetic nephropathy, other microvascular complications (diabetic neuropathy, diabetic eye complications, diabetic foot/peripheral angiopathy), and death.

Treatment Recommendation Models

The treatment recommendation models were based on the patient's clinical information from visits as input to recommend three types of treatments as output: antiglycemic, antihypertensive, and lipid-lowering medications. The input clinical information of a patient contains demographic information, lab data, physical measurements, medical history, and prescriptions currently in use. We utilized three models to recommend the three types of medications, and then combined the output of the three models into a comprehensive treatment recommendation.

Figure 1 illustrates the treatment recommendation approach by combining a knowledge-driven model and a data-driven model. The knowledge-driven model was developed based on the clinical guidelines and expert experiences on managing T2DM [29-34]. For the data-driven model, RL was used to learn the policy of treatment recommendation from real-world data that optimizes a predefined long-term goal via trial and error [35-37]. When integrating these two types of models, the knowledge-driven model was first applied to select the candidate

XSL•FO
RenderX

medications, and the data-driven model was used to rank the candidates by the expected clinical outcomes.

RL-Based Framework of the Data-Driven Model

We trained the RL model on a set of time-varying data consisting of state s_t (clinical data of the current visit), action a_t (treatment), and reward r_t score (based on clinical outcome). The ultimate goal of RL is to learn a policy π, which for any given state, one can select the action that maximizes cumulative future rewards.

The DQN [38] is a type of RL method that has been recently utilized to solve clinical decision problems with continuous state variables [13,39,40]. Referring to the previous DQN work of sepsis treatment in the intensive care unit [13], we applied deep neural networks to calculate the action-value function Q that estimates the cumulative rewards for each treatment action at the current visit state. To train the DQN model, two neural networks with the same architecture were used: an evaluation network $Q(\cdot)$ and a target network . The evaluation network was used to obtain optimal action $\max_a Q(s_t, a, \theta)$ and was trained by the loss function $L(Q', Q)$. The target network was used to estimate the expected action-value Q' to calculate the loss function L and updated its parameters  by slowly tracking the parameters of the evaluation network θ every training iteration:  $\leftarrow\tau\cdot\theta+\theta(1-\tau)\cdot$  with update parameter $\tau<1$. The loss function $L(Q', Q)$ is defined in Equation (1):

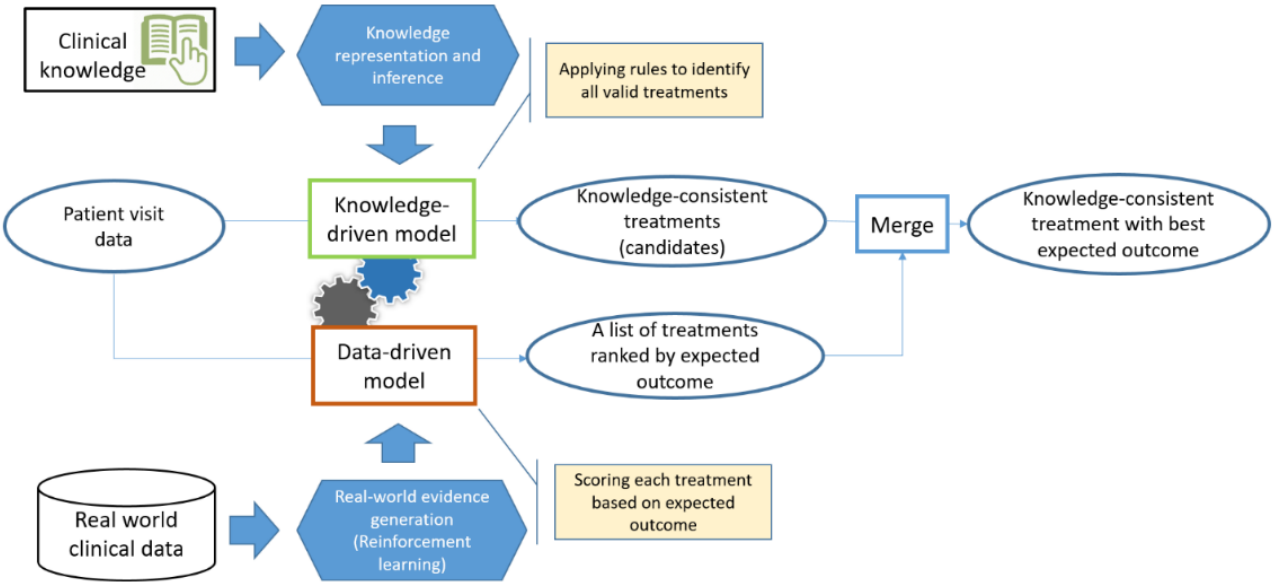


where the expected Q' is:



and r_{reg} is the maximum reward of all r_t that is used in the regularization term to penalize an inexpertly large Q value. Here, we used a double-DQN [36] architecture that calculates  in Equation (2) through action from the evaluation network $a'=argmax_a\ [Q(s_{t+1}, a, \theta)]$ instead of max . Double-DQN leads to a more stable learning target and low-variance action-value estimates. Moreover, we used dueling-DQN [37] that adds a dueling architecture in the network to separate the output of the last hidden layer into two streams to learn state values and state-independent action advantages, respectively. We also used a prioritized experience replay [41] method to speed up the training approach. Each training batch was sampled from the training data according to the importance, which was measured by the samples' temporal-difference error. The complete training procedure of our DQN model can be found in Algorithm S1 in Multimedia Appendix 1.

Figure 1. Treatment recommendation model of the "Knowledge + Data" two-wheel-drive method.



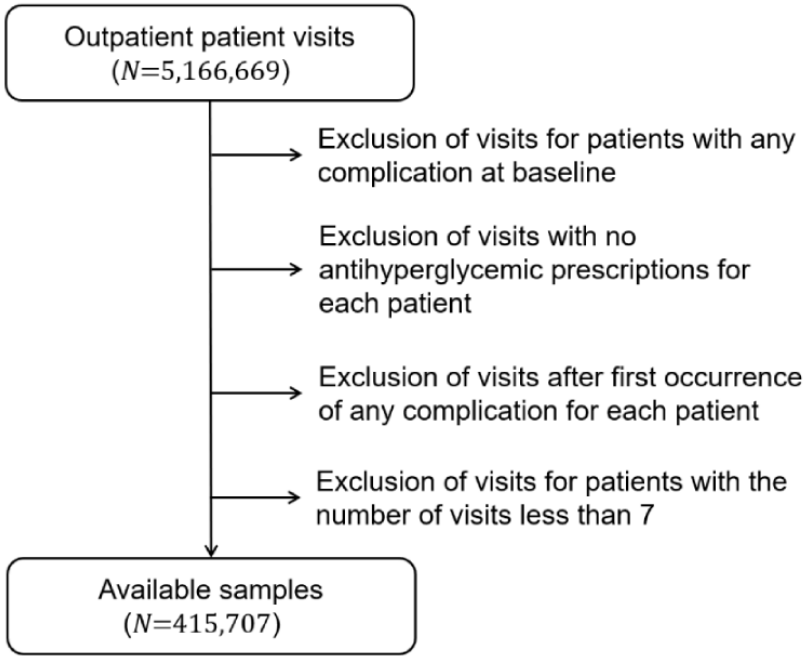
Details of DQN Implementation on the Dataset

To present the specific design of the RL-based data-driven model, we take antiglycemic treatment as an example to explain how we trained and applied the DQN in our models.

The given dataset makes up 80% of the entire dataset, which contained 152,527 patients and 5,166,669 outpatient visits. To define outpatient visit samples from the EMRs, we used lab test results (or physical measurements) within a certain time interval before the visit as state variables of its corresponding sample. The complications that occurred before the current outpatient visit were treated as medical history. The criteria used in the generation of a given dataset are shown in Figure 2. Subsequently, the patients in the given dataset were randomly divided into a training set and validation set at a ratio of 8:2.

XSL•FO
RenderX

Figure 2. Training set and validation set generation criteria for antiglycemic treatment recommendation.



The features that were selected to define the states for a patient visit included demographics, medical history, disease risks, previous drugs, lab data, and physical measurements. The detailed state information is presented in Table S1 in Multimedia Appendix 1. Among these features, continuous variables were normalized into a common scale, whereas binary variables were represented using 0 or 1. Other categorical variables were converted into multiple binary variables using one-hot encoding. Finally, we obtained a 49-dimension state vector.

For patients with T2DM, antiglycemic medications are usually prescribed based on the currently used drugs [29,32,33]. Thus, to simplify the action space of the DQN, we defined an action of a visit based on medication changes from the previous prescription. The prescription changes included drug changes at the drug class level and dosage changes of some common drugs. In terms of dosage changes, the increase or decrease in dosage of the three most frequently used drugs, namely metformin, basal insulin, and premixed insulin, was considered. Table 1 lists the actions for antiglycemic treatment used in the DQN model. The most common medication adjustment options in the actions were: changing the dosage of a drug, adding an oral antidiabetic drug (OAD), and changing to insulin treatment. Among the action options, “No prescription change” indicates use of the same drugs and dosages as the previous prescription, and “Using xxx insulin” means changing to the specific insulin or insulin combinations.

Table 1. Actions of the deep Q network for antiglycemic treatment.

Number	Action
0	No prescription change
1	Increase drug dosage
2	Decrease drug dosage
3	Adding alpha-glucosidase inhibitor
4	Adding dipeptidyl peptidase-4 inhibitor
5	Adding metformin
6	Adding sodium glucose cotransporter-2 inhibitor
7	Adding sulfonylurea
8	Adding thiazolidinedione
9	Adding glucagon-like peptide-1 receptor agonist
10	Using basal insulin
11	Using premix insulin
12	Using basal and prandial insulins

XSL•FO
RenderX

The reward function is usually defined to quantify the effectiveness of the action at each time step. In the antiglycemic DQN, for a patient visit at time *t*, we defined the reward function r_t as shown in Equation (3):

$$\text{reward}_{gly}=a \cdot \text{sgn}(7 - s_{t+1}^{HbA1c}) + (-b) \cdot s_{t+1}^{Hypo} + (-c) \cdot s_{t+1}^{Final} \cdot s_{t+1}^{CX} \quad \textbf{(3)}$$

where



(4)

In Equation 3, s_{t+1}^{HbA1c} (%) is the HbA_{1c} of time t+1, s_{t+1}^{Hypo}, is a binary (1,0) variable representing whether hypoglycemia occurs before time *t*+1, s_{t+1}^{Final} is a binary (1,0) variable representing whether time *t*+1 is the final visit of the patient, and $s_{t+1}^{CX} \in \{-1,1\}$ indicates whether complications or death occur at time *t*+1, where –1 represents "No" and 1 represents "Yes."

The concept underlying the reward function is to give a positive reward when (1) the HbA_{1c} after 3-6 months reaches the control target (less than 7%), and (2) no complications or death occurred until the last visit of a patient in the next 6 years. A negative reward (ie, penalty) is given when (1) the HbA_{1c} after 3-6 months is not well controlled, (2) a hypoglycemia event occurs in the next 6 months, and (3) a complication or death occurs after the current visit. Based on the importance of these outcomes, we set the coefficients in Equation (1) as follows: *a*=1, *b*=2, *c*=4. For an intermediate visit of a patient, the DQN model is trained to optimize the cumulative reward, which is equal to the current reward plus the next visit's expected cumulative reward multiplied by a discount factor, γ=0.9. Therefore, the DQN model is able to estimate the impact of a current action on both short-term and long-term outcomes.

The network architecture and training settings are provided below. We adopted a fully connected neural network with 2 hidden layers of 64, with 32 units for the Q networks. Each hidden layer contained batch normalization and Leaky-ReLU activation. The input layer was 49 dimensions and the output layer was 14 dimensions, which were the same as the sizes of the state vector and the action space. The learning rate η was 0.001, the batch size was 256, and the target network update parameter τ was set to 0.01. For regulation, we set the reward threshold r_{reg}=4 and λ=0.5. We trained the DQN model for a maximum of 100,000 iterations using the Adam optimizer [42].

For antihypertensive and lipid-lowering treatments, actions and reward functions of DQNs are shown in Table 2 and Table 3, respectively.

Table 2. Actions of the deep Q network for antihypertensive treatment.

Number	Action
0	No drugs
1	Using A[a]
2	Using B[b]
3	Using C[c]
4	Using D[d]
5	Using A and B
6	Using A and C
7	Using A and D
8	Using B and C
9	Using B and D
10	Using C and D
11	Using A, B, and C
12	Using A, B, and D
13	Using A, C, and D
14	Using B, C, and D
15	Using A, B, C, and D

[a]Angiotensin-converting-enzyme inhibitor or angiotensin II receptor blocker.

[b]Beta blocker.

[c]Calcium channel blocker.

[d]Diuretic.

XSL•FO
RenderX

Table 3. Actions of the deep Q network for lipid-lowering treatment.

Number	Action
0	No drugs
1	Using statin
2	Using fibrate
3	Using ezetimibe
4	Using statin and fibrate
5	Using statin and ezetimibe
6	Using fibrate and ezetimibe
7	Using statin, fibrate, and ezetimibe

In the antihypertensive DQN, for a patient visit at time *t*, we defined the reward function as in Equation (5):



s_{t+1}^{SBP}(mmHg) is the SBP of time *t*+1, s_{t+1}^{DBP} (mmHg) is the DBP of time *t*+1, and other terms are defined as in the reward function of the antiglycemic DQN.

In the lipid-lowering DQN, for a patient visit at time *t*, we defined the reward function as in Equation (7):

$$Reward_{lip} = a \cdot \mathrm{sgn}(2.6 - s_{t+1}^{LDL\text{-}c}) + (-c) \cdot s_{t+1}^{Final} \cdot s_{t+1}^{CX} \quad (7)$$

where $s_{t+1}^{LDL\text{-}c}$ (mmol/L) is the LDL-c of time *t*+1, and other terms are defined as in the reward function of the antiglycemic DQN.

Evaluation Methods

Short-Term Outcome Evaluation

Similar to previous works [27,43,44], we took model concordance as the exposure factor, which was determined by whether the actual prescription from the physician is concordant to the model-recommended medication. Thus, we partitioned the patient visits into a model-concordant group and a model-nonconcordant group. The short-term clinical outcomes were then compared between the two groups in terms of the goal-achieving rate of the key parameters, including blood glucose control, blood pressure control, blood lipids control, and hypoglycemia events.

For each short-term outcome, we (1) followed the first 2 exclusion steps in the dataset generation process shown in Figure 2 for the corresponding treatment type (antiglycemic, antihypertensive, or lipid-lowering treatment) and (2) excluded visits without the short-term outcome information. Thus, one patient may contribute different patient-visit samples for evaluation of these outcomes. We applied the corresponding treatment recommendation model onto these patient-visit samples to generate model-recommended medications, and partitioned the patient-visit samples into the model-concordant group and model-nonconcordant group according to the physicians' prescription. Short-term outcomes were compared between the model-concordant treatment and the model-nonconcordant treatment with significance of differences assessed via a χ^2 test.

Furthermore, we combined stratification, PS methods, and multivariate regression to adjust confounders. We first stratified the patient visits by the confounder (eg, current HbA_{1c}) that was most strongly correlated to the clinical outcome. We then performed the PS inverse probability weighting method [45,46] to adjust multiple confounders for both overall samples and stratified samples, since PS methods have been increasingly used to control confounders [47,48] in observational studies, especially for causal effect analysis. Finally, weighted multivariate logistic regression was applied to adjust for residual imbalances that might exist after PS modeling, and the adjusted odds ratios (ORs) and 95% CIs in multivariate regression were used to reveal the relationship between model concordance and short-term outcome.

Long-Term Outcome Evaluation

At the patient level, model-concordant rates were included as independent variables to evaluate the performance of the combined treatments with antiglycemic, antihypertensive, and lipid-lowering medications. The model-concordant rate was calculated for each patient by dividing the number of model-concordant visits by the total number of visits. We defined the model-concordant rate to quantify the extent to which each patient complied with the model recommendations.

We followed a similar process as shown in Figure 2 to generate three test datasets. The difference was that patient visits after the first occurrence time of the corresponding complication were removed instead of the first occurrence time of the earliest occurring complication. To describe the relationship between the patient's model-concordant rate and the occurrence rate of the long-term outcome, we present illustration curves for each type of treatment and each kind of long-term outcome, and calculated the slopes by fitting the curves with a linear function for exploring the trend. Multivariate logistic regression was further used to investigate the associations between the three types of comprehensive treatments and long-term outcomes, where patients with the three kinds of model-concordant rates were included as test samples. In the multivariate regression, the three concordant rates were included as independent variables, and the predicted risk score at baseline (representing the effects of multiple risk factors on the occurrence of a complication or death) was included as a covariate for

XSL•FO
RenderX

confounder adjustment. Coefficients and *P* values are reported for both independent variables and covariates.

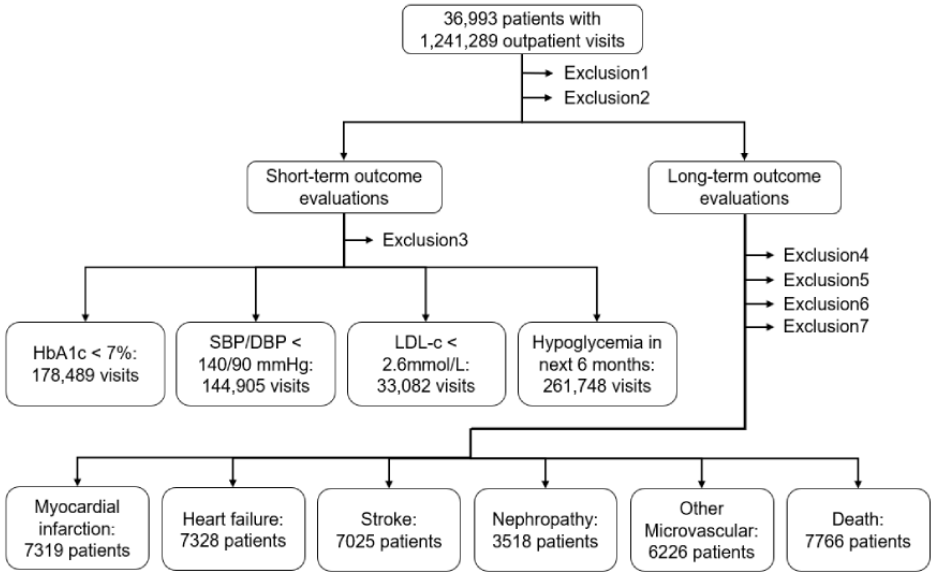
Results

Patient Characteristics

Of the 36,993 patients in the test data, 18,878 were men (51%). With respect to ethnicity, the majority of the patients were Chinese (69%), followed by Malay (15%) and Indian (11%). By 2019, the median age was 67 years (IQR 59-76) and the median duration of diabetes was 10 years (IQR 6-16).

An overview of the short-term and long-term outcome evaluation cohorts is shown in Figure 3. Further information is provided in the following two subsections.

Figure 3. Overview of the exclusion criteria and the number of visits or patients in each evaluation cohort. SBP: systolic blood pressure; DSB: diastolic blood pressure; HbA_{1c}: glycated hemoglobin A_{1c}; LDL-c: low-density lipoprotein cholesterol.



Exclusion1: Exclusion of visits for patients with any complication at baseline
Exclusion2: Exclusion of visits with no corresponding prescription (antiglycemic, antihypertensive or lipid-lowering)
Exclusion3: Exclusion of visits with no corresponding short-term outcomes
Exclusion4: Exclusion of visits after the first occurrence of corresponding complications
Exclusion5: Exclusion of visits without the three types of treatment recommendation
Exclusion6: Exclusion of patients with less than 7 visits
Exclusion7: Exclusion of patients without the corresponding risk score

Short-Term Outcomes at the Patient-Visit Level

To evaluate the short-term outcomes, we took model concordance as the exposure variable to evaluate the effect of our treatment recommendation model at the patient-visit level. We generated the test datasets separately for different short-term outcomes, namely the percentages of patient visits with well-controlled parameters (HbA_{1c}<7% [53 mmol/mol], SBP/DBP<140/90 mmHg, LDL-c<2.6 mmol/L [100 mg/dl]) after 3-6 months of therapy. For a hypoglycemia event, the occurrence rate in the following 6 months was compared between two groups. In short-term evaluation, potential confounding factors were adjusted by stratification, the PS weighting method, and multivariate regression, such as age, gender, and ethnicity.

Specifically, the model concordance was defined at the level of the standard drug class, since the treatment recommendation models output the standard drug class rather than the specific brand name. For example, if the model recommends alpha-glucosidase inhibitors (AGI), the patient visit with a prescription for Acarbose (a type of drug belonging to the drug class AGI) is model-concordant. Only the top-ranking recommended medication for antiglycemic, antihypertensive, and lipid-lowering therapy was considered for evaluation.

After meeting all exclusion criteria, a total of 178,489 visits were included to evaluate the short-term clinical outcomes of HbA_{1c} control. Of the total samples, 78,670 patient visits (44.08%) were model-concordant and 99,819 (55.92%) were nonconcordant. The characteristics of the model-nonconcordant and model-concordant groups in the test data are shown in Table 4 for short-term blood glucose control evaluation, and patient characteristics for other short-term outcomes (eg, blood pressure control) are shown in Tables S2-S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Table 4. Characteristics of the glycated hemoglobin A_{1c} (HbA_{1c}) cohort.

Variables	Model-nonconcordant group (n=99,819)	Model-concordant group (n=78,670)	*P* value[a]
Age (years), mean (SD)	64.17 (12.23)	64.61 (11.68)	<.001
Gender (female), n (%)	48,450 (48.5)	39,386 (50.1)	<.001
Ethnicity, n (%)			
Chinese	69,019 (69.1)	56,622 (72.0)	<.001
Indian	10,658 (10.7)	7,416 (9.4)	<.001
Malay	15,961 (16.0)	11,701 (14.9)	<.001
Smoker/exsmoker, n (%)	10,008 (10.0)	7644 (9.7)	.03
Duration of diabetes (years), mean (SD)	12.14 (8.11)	10.01 (7.52)	<.001
HbA_{1c} (%), mean (SD)	8.08 (1.39)	6.94 (1.15)	<.001
SBP[b] (mmHg), mean (SD)	133.06 (16.77)	131.71 (16.47)	<.001
DBP[c] (mmHg), mean (SD)	69.72 (9.40)	69.33 (9.35)	<.001
LDL-c[d] (mmol/L) mean (SD)	2.27 (0.76)	2.19 (0.70)	<.001
TG[e] (mmol/L), mean (SD)	1.61 (0.95)	1.46 (0.79)	<.001
BMI (kg/m^2), mean (SD)	26.75 (5.79)	26.54 (5.95)	<.001
eGFR[f] ($mL \cdot min^{-1} \cdot 1.73m^{-2}$), mean (SD)	80.56 (33.84)	84.14 (30.65)	<.001
Hypertension, n (%)	92,235 (92.4)	72,066 (91.6)	<.001
Hypercholesterolemia, n (%)	97,506 (97.7)	76,385 (97.1)	<.001
Macrovascular complications, n (%)	26,347 (26.4)	19,604 (24.9)	<.001
Microvascular complications, n (%)	44,904 (45.0)	29,614 (37.6)	<.001

[a]Based on a *t* test or χ^2 test.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]LDL-c: low-density lipoprotein cholesterol.

[e]TG: triglycerides.

[f]eGFR: estimated glomerular filtration rate.

The evaluation results of short-term outcomes, including HbA_{1c}, SBP/DBP, LDL-c control, and hypoglycemia event, were based on test samples with corresponding outcome data during the follow-up period. After confounder adjustment for patient characteristics, the model-concordant treatments were associated with good blood glucose control, good blood pressure control, and good blood lipid control compared with model-nonconcordant treatments. There was no significant difference in the occurrences of hypoglycemia events between model-concordant treatments and model-nonconcordant treatments (Table 5).

We further stratified the patient visits by the confounder that was most strongly correlated to the clinical outcome. For the outcome of glucose control (ie, HbA_{1c} after 3-6 months), the current HbA_{1c} of the patient was used to stratify the patient visits into three groups of low (<7%), medium (7-9%), and high (>9%) levels. The short-term evaluation was performed on each group separately (Table 6), showing that model-concordant treatments were associated with improved short-term HbA_{1c} outcomes (ie, higher HbA_{1c} goal-achieving rate) in each group.

XSL•FO
RenderX

Table 5. Short-term clinical outcomes in the model-concordant and model-nonconcordant groups.

Short-term outcomes	Samples, n	Before adjustment		After propensity score weighting adjustment		
		Incidence, n (%)	*P* value	OR[a]	95% CI	*P* value
Antiglycemic treatment[b] (HbA_{1c})[c] <7%)			<.001	1.73	1.69-1.76	<.001
Concordant	78,670	48,263 (61.35)				
Nonconcordant	99,819	21,507 (21.55)				
Antihypertensive treatment[d] (SBP[e]/DBP[f]<140/90 mmHg)			<.001	1.26	1.23-1.29	<.001
Concordant	80,868	62,058 (76.74)				
Nonconcordant	64,037	35,327 (55.17)				
Lipid-lowering treatment[g] (LDL-c[h]<2.6 mmol/L)			<.001	1.28	1.22-1.35	<.001
Concordant	14,985	10,702 (71.42)				
Nonconcordant	18,097	10,028 (55.41)				
Antiglycemic treatment[i] (Hypoglycemia in next 6 months)			<.001	0.97	0.91-1.02	0.22
Concordant	113,343	1497 (1.32)				
Nonconcordant	148,405	3009 (2.03)				

[a]OR: odds ratio.

[b]Confounders considered: age, gender, ethnicity, smoking, duration of diabetes, HbA_{1c}, SBP/DBP, LDL-c, triglycerides (TG), BMI, estimated glomerular filtration rate (eGFR), hypertension, hypercholesterolemia, macrovascular complication, microvascular complication, hypoglycemia history.

[c]HbA_{1c}: glycated hemoglobin A_{1c}.

[d]Confounders considered: age, gender, ethnicity, smoking, duration of diabetes, HbA_{1c}, SBP/DBP, LDL-c, TG, BMI, eGFR, hypercholesterolemia, myocardial infarction, unstable angina, heart failure, stroke, nephropathy.

[e]SBP: systolic blood pressure.

[f]DBP: diastolic blood pressure.

[g]Confounders considered: age, gender, ethnicity, smoking, duration of diabetes, HbA_{1c}, SBP/DBP, LDL-c, TG, BMI, eGFR, alanine transaminase, macrovascular complication, nephropathy.

[h]LDL-c: low-density lipoprotein cholesterol.

[i]Confounders considered: age, gender, ethnicity, smoking, duration of diabetes, HbA_{1c}, SBP, LDL-c, BMI, serum creatinine, hypertension, atrial fibrillation, macrovascular complication, microvascular complication.

XSL•FO
RenderX

Table 6. Short-term outcome for antiglycemic treatment based on current glycated hemoglobin A_{1c} (HbA_{1c})

Current HbA_{1c} level	Samples, n	Before adjustment		After propensity score weighting adjustment		
		Incidence, n (%)	*P* value	OR[b]	95% CI	*P* value
Low (<7%)			<.001	1.79	1.69-1.89	<.001
Concordant	60,428	43,836 (72.54)				
Nonconcordant	5959	4034 (67.7)				
Medium (7-9%)			<.001	1.76	1.69-1.83	<.001
Concordant	14,025	3816 (27.21)				
Nonconcordant	76,423	16,365 (21.41)				
High (>9%)			<.001	1.83	1.63-2.05	<.001
Concordant	4217	611 (14.49)				
Nonconcordant	17,437	1108 (6.35)				

[a]Confounders considered: age, gender, ethnicity, smoking, duration of diabetes, HbA_{1c}, systolic blood pressure/diastolic blood pressure, low-density lipoprotein cholesterol, triglycerides, BMI, estimated glomerular filtration rate, hypertension, hypercholesterolemia, macrovascular complication, microvascular complication, hypoglycemia history.

[b]OR: odds ratio.

Long-Term Outcomes at the Patient Level

Figure 4 illustrates the relationship between the patient's model-concordant rate and the occurrence rate of long-term clinical outcomes for all patients with respect to antiglycemic, antihypertensive, and lipid-lowering therapy. Specifically, the patients were divided into different groups according to the patient's model-concordant rate (eg, every 20% as a group), and the occurrence rate of complications or death in each group was computed. In general, the curves show a downward trend. In other words, there is a negative correlation between the model-concordant rate and the occurrence rate of complications or death; the higher the patient's model-concordant rate, the lower the occurrence rate of complications or death. Table 7 shows the slope of each curve by fitting to a straight line using all data points, which indicates the extent of the downward trend. In addition, the number of patients in each long-term outcome evaluation cohort is shown in Table S5 in Multimedia Appendix 1.

Furthermore, for combined treatments with antiglycemic, antihypertensive, and lipid-lowering drugs, we evaluated if the patient's model-concordant rate for the three types of treatments had a positive impact on the reduction of the complication or death risk by multivariate regression. Only patients with all three model-concordant rates of antiglycemic, antihypertensive, and lipid-lowering treatment were included in the multivariate regression model. Table 8 shows the multivariate regression results of each long-term outcome with confounder adjustment for the corresponding risk score generated by the prediction model. All of the prediction models, based on XGBoost, outperformed the clinical baseline models [49-53] and demonstrated good prediction capability, with an area under the receiver operating characteristic curve ranging from .711 to .874. The model-concordant rate for antiglycemic treatment had a negative correlation with the occurrence of major complications and death, with coefficients ranging from –1.12 to –.33. A similar result was found for the model-concordant rate for antihypertensive treatment (coefficient range –1.44 to –.40) and the model-concordant rate for lipid-lowering treatment (coefficient range –1.17 to –.52). All of these coefficients were significant (P<.05), except for the coefficients of antiglycemic treatment and antihypertensive treatment in the evaluation of stroke outcome. This implies that the patients whose treatments were more concordant with the model recommendation were more likely to be associated with a lower risk of diabetes complications (including both macrovascular and microvascular complications) and death. All of the coefficients for the risk score had positive values, which also validated the soundness of the risk prediction models. In addition, the number of patients with or without the corresponding long-term outcomes in multivariate regression are shown in Table S6 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 4. Relationship between patient's model-concordant rate and the occurrence rate of long-term outcomes for all patients with respect to antiglycemic, antihypertensive, and lipid-lowering treatment, respectively.

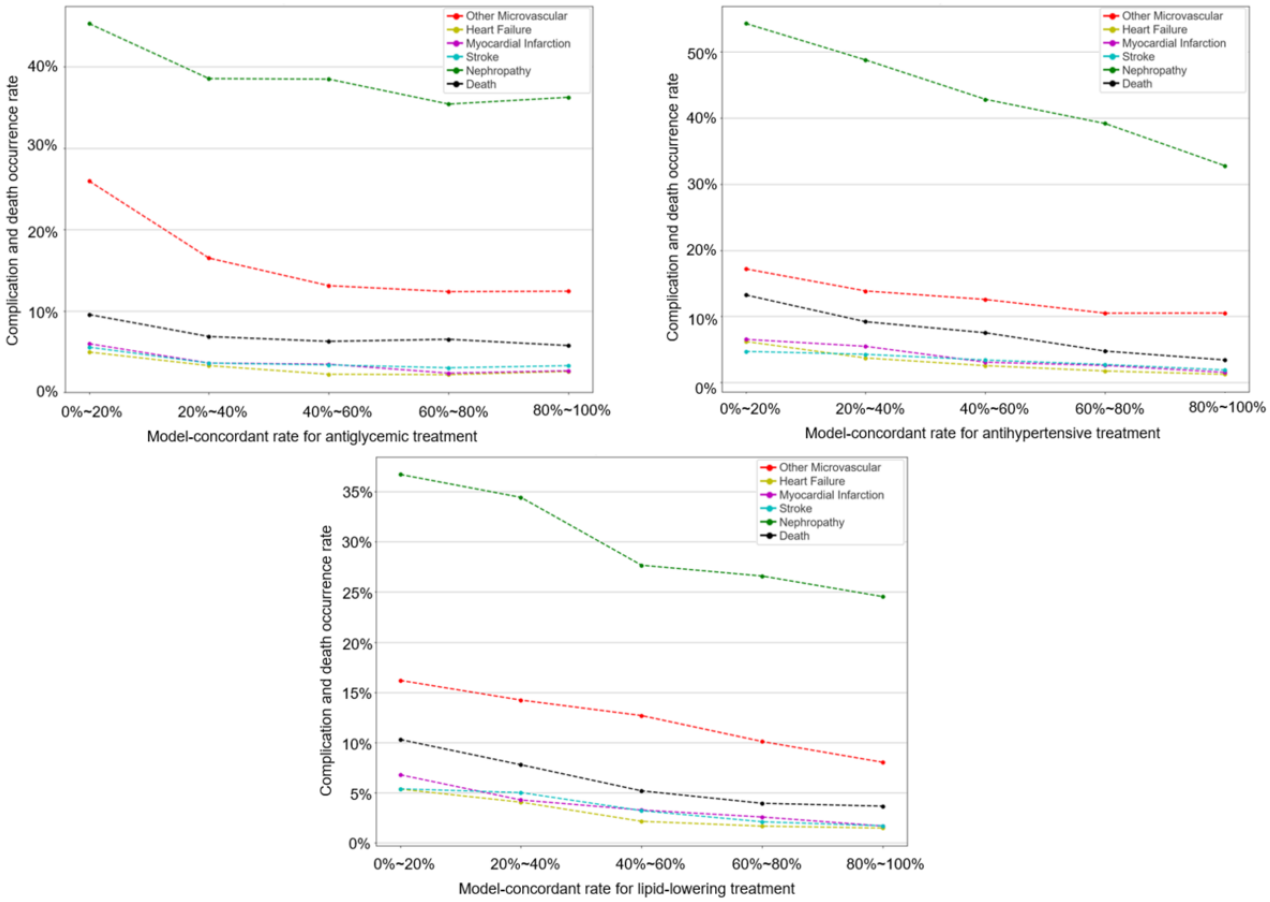


Table 7. Slopes of patient-level long-term outcome evaluation curves in Figure 4.

Curves fit	Antiglycemic treatment	Antihypertensive treatment	Lipid-lowering treatment
Myocardial infarction	–0.0384	–0.0617	–0.0581
Heart failure	–0.0312	–0.057	–0.0502
Stroke	–0.0261	–0.0389	–0.0535
Nephropathy	–0.1098	–0.2601	–0.1670
Other microvascular	–0.1584	–0.0827	–0.1007
Death	–0.0419	–0.1240	–0.0881

Table 8. Multivariate regression results for long-term outcomes.

Long-term outcome	Antiglycemic treatment model concordance rate		Antihypertensive treatment model concordance rate		Lipid-lowering treatment model concordance rate		Risk score (%)	
	Coefficient (β)	*P* value	Coefficient (β)	*P* value	Coefficient (β)	*P* value	Coefficient (β)	*P* value
Myocardial infarction	–1.1150	<.001	–.8018	<.001	–1.0065	<.001	.0998	<.001
Heart failure	–.6294	.04	–1.4414	<.001	–1.0416	<.001	.0653	<.001
Stroke	–.3288	.24	–.4871	.06	–1.1715	<.001	.1210	<.001
Nephropathy	–.5667	<.001	–1.2648	<.001	–.5182	<.001	.0072	<.001
Other Microvascular	–.5382	.001	–.3956	.004	–.6296	<.001	.0552	<.001
Death	–.4922	.02	–.8872	<.001	–.9835	<.001	.0527	<.001

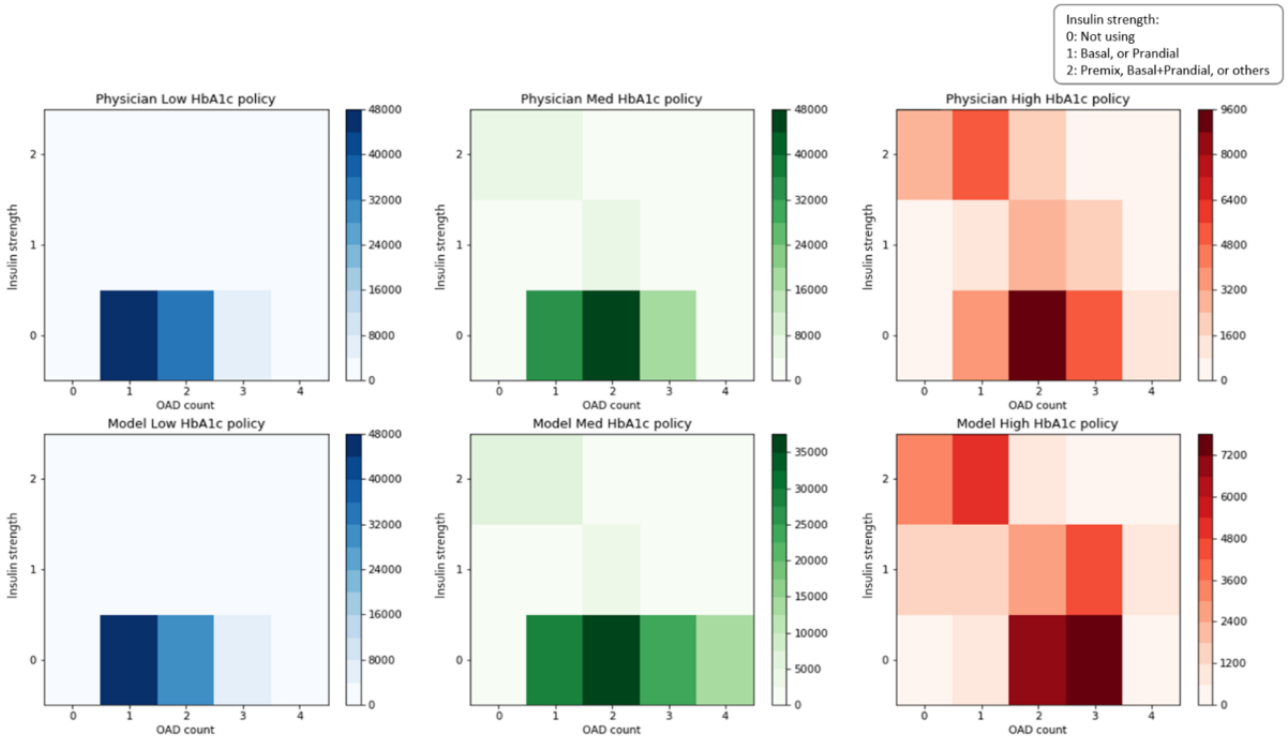
XSL•FO
RenderX

Medication Pattern

In all patient visits, the percentages of model-concordant visits were 43.30%, 51.25%, and 58.93% for antiglycemic, antihypertensive, and lipid-lowering therapy, respectively. To analyze the distribution characteristics for model concordance, we compared the antiglycemic medication patterns between physicians' prescriptions and model recommendations.

Figure 5 depicts the medication patterns used by physicians (top panels) and the model (bottom panels) for the three different groups. All patient-visit samples were categorized into three groups based on the current HbA_{1c} as follows: low (<7%), medium (7-9%), and high (>9%). For each group, the medication patterns of physicians' prescriptions and model recommendations are visualized by 2D histograms, in which the x-axis represents the number of OADs and the y-axis represents the intensity of insulin (a value of 0 indicates no insulin used, 1 indicates single use of basal insulin or prandial insulin, and 2 indicates a premix or combination use of basal and prandial insulin, or others). The color indicates the usage number of corresponding medication patterns. First, Figure 5 shows that medication patterns of model recommendations are consistent with clinical knowledge, as most patients in the low group were prescribed with a single OAD, whereas patients in the medium and high groups showed increased use of multi-OAD, insulin, and insulin plus OAD. Second, the medication patterns of model recommendations are visually similar to those of physicians' prescriptions in the low group, whereas in the medium and high groups, the patterns of model recommendations are more vigorous than those of physicians' prescriptions. This indicates that the model learns from the data, showing that active adjustment of the medication for the patients in medium and high groups may be associated with a better clinical outcome.

Figure 5. Medication pattern comparison between physicians' prescriptions and model recommendations. HbA_{1c}: glycated hemoglobin A_{1c}; OAD: oral antidiabetic drug.



Discussion

In this work, we built our treatment recommendation model based on 80% of the data in SingHealth Diabetes Registry and evaluated its effectiveness with the remaining 20% of data. The strengths of this retrospective study are two-fold. First, the diabetes registry used for model building and evaluation is of good quality. It consists of the medical records for a large patient population, covers a long-term span of 6 years, and includes different types of diabetes complications (macrovascular and microvascular). Second, the methods used for the evaluation are comprehensive. For the treatment recommendation, we considered the concordance of three types of treatments (antiglycemic, antihypertensive, and lipid-lowering treatment) and evaluated two types of outcomes, namely the control of key indicators in the short term and the occurrences of diabetes complications in the long term.

The treatments are recommended by a combination of a knowledge-driven model and a data-driven model. For the knowledge-driven model, we incorporated renal dosing and contraindications for specific medications so as to align with standards of care. For example, the model will not recommend an increase in the metformin dose in subjects with an estimated glomerular filtration rate (eGFR) <45 $mL \cdot min^{-1} \cdot 1.73\ m^{-2}$ and will recommend the discontinuation of metformin if the eGFR is below 30 $mL \cdot min^{-1} \cdot 1.73\ m^{-2}$ [29]. For the data-driven model, we ranked the candidates by the expected clinical outcomes.

In our study, the antiglycemic medications recommended by our model were the same as the actual prescriptions of the

XSL•FO
RenderX

physicians in 43.30% of patient visits. The percentages of model-concordant visits for antihypertensive medications and lipid-lowering medications were 51.25% and 58.93%, respectively. For the treatment recommendation evaluation, patients with more model-concordant treatments had better control of blood glucose (OR 1.73, 95% CI 1.69-1.76), blood pressure (OR 1.26, 95% CI 1.23-1.29), and blood lipids (OR 1.28, 95% CI 1.22-1.35), as well as a lower risk of diabetes complications (coefficients of regression ranging from –1.44 to –.33). In addition, there was no significant difference (OR 0.97, 95% CI 0.91-1.02) on the risk of hypoglycemia events between model-concordant treatments and model-nonconcordant treatments. These evaluation results suggested that the treatment recommendation model has good potential to guide physicians in prescribing medications that could help to achieve better clinical outcomes.

Our study has some limitations. First, the algorithm was more aggressive in recommending complex treatment regimens than the actual physicians' prescriptions, especially in the medium and high HbA_{1c} groups. The dataset is built based on the EMRs of patients with T2DM. Some information that can influence the choice of a physician's prescription may be missing in the data. For example, a physician may default to repeating a previous prescription because of the patient's reluctance to change medications. Such patient preference will not be recorded in the data and hence not used in the treatment recommendation model. This is reflective of "human bias" for less complex treatment regimens in real-world clinical practice. Second, selection bias may exist in this study. For example, when evaluating the long-term clinical outcome, we selected the patients with a number of visits greater than a threshold. As such, newly added patients in 2018 were hardly selected. Third, the unified therapeutic targets were used in this study without considering personalized control targets for individual patients. For example, the control goal of blood glucose was set to HbA_{1c}<7% in this study. However, for elderly patients and patients with recurrent hypoglycemia, the HbA_{1c} goal could be less strict. Fourth, the hypoglycemia episodes that did not end up requiring admissions may be reported by the patients but are seldom coded in the EMRs as a blood test or diagnosis that the algorithm can identify [54]. Thus, our analysis only considered severe hypoglycemia events with hospital admissions. Finally, although we performed the PS weighting method and multivariate regression analysis to control for differences in demographic and clinical conditions when evaluating the association between model concordance and clinical outcomes, a conclusion regarding the causal effect of model concordance cannot be made based on the observed association due to the limitations of a retrospective study.

In future work, the treatment recommendation model can be further evaluated in a prospective study by piloting an interactive treatment recommendation system in a real-world clinical practice. Finally, the knowledge-driven and data-driven models need to be optimized regularly to make use of newly collected EMR data, and to incorporate the latest clinical guidelines and new classes of drugs.

Acknowledgments

This research is supported by A*STAR under its Industry Alignment Fund (Pre-Positioning) (Award reference number: H19/01/a0/023 – Diabetes Clinic of the Future).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data: Algorithm S1, Tables S1-S6.
[DOCX File , 36 KB - jmir_v23i7e27858_app1.docx]

References

1. Global report on diabetes. World Health Organization. 2016. URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [accessed 2021-01-20]
2. Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. Pharmacoeconomics 2015 Aug;33(8):811-831 [FREE Full text] [doi: 10.1007/s40273-015-0268-9] [Medline: 25787932]
3. Vinik AI, Vinik E. Prevention of the complications of diabetes. Am J Manag Care 2003 Mar;9(3 Suppl):S63-80; quiz S81 [FREE Full text] [Medline: 12653455]
4. Merino J, Florez JC. Precision medicine in diabetes: an opportunity for clinical translation. Ann N Y Acad Sci 2018 Jan;1411(1):140-152 [FREE Full text] [doi: 10.1111/nyas.13588] [Medline: 29377200]
5. Sutton R, Barto A. Reinforcement Learning: An Introduction. 1st edition. Cambridge, MA: MIT Press; 1998.
6. Tseng H, Luo Y, Cui S, Chien J, Ten Haken RK, Naqa IE. Deep reinforcement learning for automated radiation adaptation in lung cancer. Med Phys 2017 Dec;44(12):6690-6705 [FREE Full text] [doi: 10.1002/mp.12625] [Medline: 29034482]
7. Jalalimanesh A, Shahabi Haghighi H, Ahmadi A, Soltani M. Simulation-based optimization of radiotherapy: Agent-based modeling and reinforcement learning. Math Comput Simul 2017 Mar;133:235-248. [doi: 10.1016/j.matcom.2016.05.008]

XSL•FO
RenderX

8. Jalalimanesh A, Haghighi HS, Ahmadi A, Hejazian H, Soltani M. Multi-objective optimization of radiotherapy: distributed Q-learning and agent-based simulation. J Exp Theor Artif Intell 2017 Mar 05;29(5):1071-1086. [doi: 10.1080/0952813x.2017.1292319]
9. Liu Z, Ji L, Jiang X, Zhao W, Liao X, Zhao T, et al. A deep reinforcement learning approach for type 2 diabetes mellitus treatment. 2020 Presented at: 2020 IEEE International Conference on Healthcare Informatics (ICHI); November 30-December 3, 2020; Oldenburg, Germany p. 1-9. [doi: 10.1109/ichi48887.2020.9374313]
10. Zuo L, Du X, Zhao W, Jiang C, Xia S, He L, et al. Improving anticoagulant treatment strategies of atrial fibrillation using reinforcement learning. 2020 Presented at: AMIA 2020 Annual Symposium; November 14-18, 2020; Los Angeles, USA URL: https://knowledge.amia.org/72332-amia-1.4602255/t003-1.4606204/t003-1.4606205/3410444-1.4606209/3410444-1.4606210?qr=1
11. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
12. Raghu A, Komorowski M, Celi LA, Szolovits P, Ghassemi M. Continuous state-space models for optimal sepsis treatment - a deep reinforcement learning approach. 2017 Presented at: Machine Learning for Healthcare Conference; August 18-19, 2017; Northeastern University, USA p. 147-163.
13. Raghu A, Komorowski M, Ahmed I, Celi LA, Szolovits P, Ghassemi M. Deep reinforcement learning for sepsis treatment. arXiv. 2017 Nov. URL: https://arxiv.org/abs/1711.09602 [accessed 2021-01-20]
14. Nemati S, Ghassemi MM, Clifford GD. Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:2978-2981. [doi: 10.1109/EMBC.2016.7591355] [Medline: 28268938]
15. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]
16. Forman EM, Kerrigan SG, Butryn ML, Juarascio AS, Manasse SM, Ontañón S, et al. Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss? J Behav Med 2019 Apr;42(2):276-290 [FREE Full text] [doi: 10.1007/s10865-018-9964-1] [Medline: 30145623]
17. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. arXiv. 2014. URL: https://arxiv.org/abs/1406.2661 [accessed 2021-03-16]
18. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018 Apr 27;97(4):380-387. [doi: 10.1111/aogs.13309] [Medline: 29377058]
19. Neugebauer R, Fireman B, Roy JA, O'Connor PJ, Selby JV. Dynamic marginal structural modeling to evaluate the comparative effectiveness of more or less aggressive treatment intensification strategies in adults with type 2 diabetes. Pharmacoepidemiol Drug Saf 2012 May;21(Suppl 2):99-113. [doi: 10.1002/pds.3253] [Medline: 22552985]
20. Marinier K, Macouillard P, de Champvallins M, Deltour N, Poulter N, Mancia G. Effectiveness of two-drug therapy versus monotherapy as initial regimen in hypertension: A propensity score-matched cohort study in the UK Clinical Practice Research Datalink. Pharmacoepidemiol Drug Saf 2019 Dec;28(12):1572-1582 [FREE Full text] [doi: 10.1002/pds.4884] [Medline: 31482621]
21. Nissen SE, Tuzcu EM, Schoenhagen P, Brown BG, Ganz P, Vogel RA, REVERSAL Investigators. Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. JAMA 2004 Mar 03;291(9):1071-1080. [doi: 10.1001/jama.291.9.1071] [Medline: 14996776]
22. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. J Am Coll Cardiol 2017 Jan 24;69(3):345-357 [FREE Full text] [doi: 10.1016/j.jacc.2016.10.060] [Medline: 28104076]
23. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health 1999;20:145-157. [doi: 10.1146/annurev.publhealth.20.1.145] [Medline: 10352854]
24. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000 Sep;11(5):550-560. [doi: 10.1097/00001648-200009000-00011] [Medline: 10955408]
25. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. Am J Epidemiol 2002 Jun 01;155(11):1045-1053. [doi: 10.1093/aje/155.11.1045] [Medline: 12034583]
26. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000 Sep;11(5):561-570. [doi: 10.1097/00001648-200009000-00012] [Medline: 10955409]
27. Chen S, Lee Y, Alas V, Greene M, Brixner D. Outcomes associated with nonconcordance to National Kidney Foundation guideline recommendations for oral antidiabetic drug treatments in patients with concomitant type 2 diabetes and chronic kidney disease. Endocr Pract 2014 Mar;20(3):221-231. [doi: 10.4158/EP13269.OR] [Medline: 24246342]
28. Lim DYZ, Chia SY, Abdul Kadir H, Mohamed Salim NN, Bee YM. Establishment of the SingHealth Diabetes Registry. Clin Epidemiol 2021 Mar;13:215-223. [doi: 10.2147/clep.s300663]

XSL•FO
RenderX

29. Goh S, Ang S, Bee Y, Chen Y, Gardner D, Ho E, et al. Ministry of Health Clinical Practice Guidelines: Diabetes Mellitus. Singapore Med J 2014 Jun 30;55(6):334-347 [FREE Full text] [doi: 10.11622/smedj.2014079] [Medline: 25017409]
30. Tay JC, Sule AA, Chew EK, Tey JS, Lau T, Lee S, et al. Ministry of Health Clinical Practice Guidelines: Hypertension. Singapore Med J 2018 Jan;59(1):17-27. [doi: 10.11622/smedj.2018007] [Medline: 29376186]
31. Tai ES, Chia BL, Bastian AC, Chua T, Ho SCW, Koh TS, et al. Ministry of Health Clinical Practice Guidelines: Lipids. Singapore Med J 2017 Mar;58(3):155-166. [doi: 10.11622/smedj.2017018] [Medline: 28361160]
32. Appropriate care guide: Initiating basal insulin in type 2 diabetes mellitus. Agency for Care Effectiveness, Ministry of Health, Republic of Singapore. 2017. URL: https://www.ace-hta.gov.sg/docs/default-source/acgs/initiating-basal-insulin-in-t2dm.pdf [accessed 2021-06-25]
33. Appropriate care guide: Oral glucose-lowering agents in type 2 diabetes mellitus - an update. Agency for Care Effectiveness, Ministry of Health, Republic of Singapore. 2017. URL: https://www.ace-hta.gov.sg/docs/default-source/acgs/oral-glucose-lowering-agents-in-t2dm-(updated-on-3-august-2017).pdf [accessed 2021-06-25]
34. Sodium-glucose co-transporter 2 (SGLT2) inhibitors for treating type 2 diabetes mellitus. Agency for Care Effectiveness, Ministry of Health, Republic of Singapore. 2017. URL: https://www.ace-hta.gov.sg/docs/default-source/drug-guidances/sodium-glucose-co-transporter-2-(sglt2)-inhibitors-for-treating-type-2-diabetes-mellitus-(updated-1-sept-2020).pdf [accessed 2021-06-25]
35. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015 Mar 26;518(7540):529-533. [doi: 10.1038/nature14236] [Medline: 25719670]
36. Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning. 2016 Presented at: Thirtieth AAAI Conference on Artificial Intelligence; February 12-17, 2016; Phoenix, AZ p. 12-17.
37. Wang Z, Schaul T, Hessel M, Hasselt H, Lanctot M, Freitas N. Dueling network architectures for deep reinforcement learning. arXiv. 2015. URL: https://arxiv.org/abs/1511.06581 [accessed 2021-01-20]
38. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing Atari with deep reinforcement learning. arXiv. 2013. URL: https://arxiv.org/abs/1312.5602 [accessed 2021-01-20]
39. Ling Y, Hasan SA, Datla V, Qadir A, Lee K, Liu J, et al. Diagnostic inferencing via improving clinical concept extraction with deep reinforcement learning: A preliminary study. 2017 Presented at: Machine Learning for Healthcare; August 18-19, 2017; Northeastern University, USA p. 18-19. [doi: 10.1007/978-1-4842-2845-6_1]
40. Tang KF, Kao HC, Chou CN, Chang EY. Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning. 2016 Presented at: NIPS Workshop on Deep Reinforcement Learning; December 5-10, 2016; Barcelona, Spain. [doi: 10.7551/mitpress/11474.003.0014]
41. Schaul T, Quan J, Antonoglou I, Silver D. Prioritized experience replay. arXiv. 2016. URL: https://arxiv.org/abs/1511.05952 [accessed 2021-01-20]
42. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv. 2014. URL: https://arxiv.org/abs/1412.6980 [accessed 2021-01-20]
43. Frei CR, Restrepo MI, Mortensen EM, Burgess DS. Impact of guideline-concordant empiric antibiotic therapy in community-acquired pneumonia. Am J Med 2006 Oct;119(10):865-871. [doi: 10.1016/j.amjmed.2006.02.014] [Medline: 17000218]
44. Gardiner FW, Nwose EU, Bwititi PT, Crockett J, Wang L. Adherence to blood pressure and glucose recommendations in chronic kidney disease hospital inpatients: Clinical inertia and patient adherence. Diabetes Metab Syndr 2018 May;12(3):291-300. [doi: 10.1016/j.dsx.2017.12.007] [Medline: 29277347]
45. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
46. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015 Dec 10;34(28):3661-3679 [FREE Full text] [doi: 10.1002/sim.6607] [Medline: 26238958]
47. Ramos R, García-Gil M, Comas-Cufí M, Quesada M, Marrugat J, Elosua R, et al. Statins for prevention of cardiovascular events in a low-risk population with low ankle brachial index. J Am Coll Cardiol 2016 Mar 16;67(6):630-640 [FREE Full text] [doi: 10.1016/j.jacc.2015.11.052] [Medline: 26868687]
48. Solomon MD, Go AS, Shilane D, Boothroyd DB, Leong TK, Kazi DS, et al. Comparative effectiveness of clopidogrel in medically managed patients with unstable angina and non-ST-segment elevation myocardial infarction. J Am Coll Cardiol 2014 Jun 03;63(21):2249-2257 [FREE Full text] [doi: 10.1016/j.jacc.2014.02.586] [Medline: 24703914]
49. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. Stroke 1994 Jan;25(1):40-43. [doi: 10.1161/01.str.25.1.40] [Medline: 8266381]
50. Chien K, Lin H, Lee B, Hsu H, Lee Y, Chen M. A prediction model for the risk of incident chronic kidney disease. Am J Med 2010 Sep;123(9):836-846. [doi: 10.1016/j.amjmed.2010.05.010] [Medline: 20800153]
51. Chow LS, Zmora R, Ma S, Seaquist ER, Schreiner PJ. Development of a model to predict 5-year risk of severe hypoglycemia in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2018;6(1):e000527 [FREE Full text] [doi: 10.1136/bmjdrc-2018-000527] [Medline: 30116541]

XSL•FO
RenderX

52. D'Agostino RB, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001 Jul 11;286(2):180-187. [doi: 10.1001/jama.286.2.180] [Medline: 11448281]
53. Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PW, Levy D. Profile for estimating risk of heart failure. Arch Intern Med 1999 Jun 14;159(11):1197-1204. [doi: 10.1001/archinte.159.11.1197] [Medline: 10371227]
54. Silbert R, Salcido-Montenegro A, Rodriguez-Gutierrez R, Katabi A, McCoy RG. Hypoglycemia among patients with type 2 diabetes: epidemiology, risk factors, and prevention strategies. Curr Diab Rep 2018 Jun 21;18(8):53 [FREE Full text] [doi: 10.1007/s11892-018-1018-0] [Medline: 29931579]

Abbreviations

AGI: alpha-glucosidase inhibitors
DBP: diastolic blood pressure
DQN: deep Q network
eGFR: estimated glomerular filtration rate
EMR: electronic medical record
HbA_{1c}: glycated hemoglobin A_{1c}
LDL-c: low-density lipoprotein cholesterol
OAD: oral anti-diabetic drug
OR: odds ratio
PS: propensity score
RCT: randomized controlled trial
RL: reinforcement learning
SBP: systolic blood pressure
SingHealth: Singapore Health Services
T2DM: type 2 diabetes mellitus

Edited by R Kukafka; submitted 09.02.21; peer-reviewed by I Hochberg, M Komorowski; comments to author 25.02.21; revised version received 30.03.21; accepted 06.05.21; published 22.07.21.

Please cite as:
Sun X, Bee YM, Lam SW, Liu Z, Zhao W, Chia SY, Abdul Kadir H, Wu JT, Ang BY, Liu N, Lei Z, Xu Z, Zhao T, Hu G, Xie G
Effective Treatment Recommendations for Type 2 Diabetes Management Using Reinforcement Learning: Treatment Recommendation Model Development and Validation
J Med Internet Res 2021;23(7):e27858
URL: https://www.jmir.org/2021/7/e27858
doi:10.2196/27858
PMID:34292166

©Xingzhi Sun, Yong Mong Bee, Shao Wei Lam, Zhuo Liu, Wei Zhao, Sing Yi Chia, Hanis Abdul Kadir, Jun Tian Wu, Boon Yew Ang, Nan Liu, Zuo Lei, Zhuoyang Xu, Tingting Zhao, Gang Hu, Guotong Xie. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Frustration With Technology and its Relation to Emotional Exhaustion Among Health Care Workers: Cross-sectional Observational Study

Daniel S Tawfik[1], MS, MD; Amrita Sinha[1], MD; Mohsen Bayati[2,3], PhD; Kathryn C Adair[4], PhD; Tait D Shanafelt[5,6], MD; J Bryan Sexton[4,7], PhD; Jochen Profit[1,8], MPH, MD

[1]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA, United States
[2]Operations, Information, and Technology, Stanford University Graduate School of Business, Stanford, CA, United States
[3]Department of Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, United States
[4]Duke Center for Healthcare Safety and Quality, Duke University Health System, Durham, NC, United States
[5]Department of Medicine, Stanford University School of Medicine, Stanford, CA, United States
[6]WellMD Center, Stanford University School of Medicine, Stanford, CA, United States
[7]Department of Psychiatry, Duke University School of Medicine, Duke University Health System, Durham, NC, United States
[8]California Perinatal Quality Care Collaborative, Palo Alto, CA, United States

Corresponding Author:
Daniel S Tawfik, MS, MD
Department of Pediatrics
Stanford University School of Medicine
770 Welch Road
Suite 435
Palo Alto, CA, 94304
United States
Phone: 1 650 723 9902
Email: dtawfik@stanford.edu

Abstract

Background: New technology adoption is common in health care, but it may elicit frustration if end users are not sufficiently considered in their design or trained in their use. These frustrations may contribute to burnout.

Objective: This study aimed to evaluate and quantify health care workers' frustration with technology and its relationship with emotional exhaustion, after controlling for measures of work-life integration that may indicate excessive job demands.

Methods: This was a cross-sectional, observational study of health care workers across 31 Michigan hospitals. We used the Safety, Communication, Operational Reliability, and Engagement (SCORE) survey to measure work-life integration and emotional exhaustion among the survey respondents. We used mixed-effects hierarchical linear regression to evaluate the relationship among frustration with technology, other components of work-life integration, and emotional exhaustion, with adjustment for unit and health care worker characteristics.

Results: Of 15,505 respondents, 5065 (32.7%) reported that they experienced frustration with technology on at least 3-5 days per week. Frustration with technology was associated with higher scores for the composite Emotional Exhaustion scale (r=0.35, P<.001) and each individual item on the Emotional Exhaustion scale (r=0.29-0.36, P<.001 for all). Each 10-point increase in the frustration with technology score was associated with a 1.2-point increase (95% CI 1.1-1.4) in emotional exhaustion (both measured on 100-point scales), after adjustment for other work-life integration items and unit and health care worker characteristics.

Conclusions: This study found that frustration with technology and several other markers of work-life integration are independently associated with emotional exhaustion among health care workers. Frustration with technology is common but not ubiquitous among health care workers, and it is one of several work-life integration factors associated with emotional exhaustion. Minimizing frustration with health care technology may be an effective approach in reducing burnout among health care workers.

(J Med Internet Res 2021;23(7):e26817) doi:10.2196/26817

XSL•FO
RenderX

KEYWORDS

frustration with technology; emotional exhaustion; professional burnout; work-life integration; biomedical technology; work-life balance; user-centered design; electronic health records; medical informatics applications; health information systems

Introduction

Technological innovation has expanded the horizons of medicine in recent decades [1], but these advances have been accompanied by increased clerical burden among physicians and other health care providers. Electronic health record (EHR) adoption has improved emphasis on quality monitoring, billing accuracy, and research, but it has often resulted in redundant documentation and inefficient workflows [2-6]. As a result, physicians have greatly increased interactions with a variety of health care information technologies (HITs), with their interactions commonly encompassing a combination of direct clinical and nonclinical goals [7]. For example, outpatient physicians now spend up to twice as much time interacting with EHRs as they do with patients [8,9].

Concurrently, symptoms of burnout, including emotional exhaustion, have risen to epidemic proportions, now affecting over 500,000 US physicians and costing the US health care system $4-5 billion annually [10-12]. Fundamentally, medicine is a human-centered endeavor, and technology can enable or distract from this focus. Unfortunately, almost 50% of physicians believe they spend an excessive amount of time on clerical tasks, and many physicians believe EHRs contribute to burnout [13-15]. High EHR task load, time spent on EHRs, and automated in-basket messages have been associated with greater emotional exhaustion [16-21]. Furthermore, the National Academy of Medicine recognizes the associations of administrative burden, technology usability, and time pressure on burnout (all of which may be attributed as EHR or HIT factors), yet the mechanisms underlying these associations are not well described [22].

One potential mechanism relating to HIT use and burnout may be frustration with technology—an emotional reaction to an obstacle preventing the fulfillment of a perceived need [23]. HIT is at risk for inducing frustration among health care workers, by virtue of its complex interfaces, frequent updates as capabilities improve, and deployment within a high-stakes environment that provides limited opportunity for dedicated training [24-26]. If frustration with technology contributes to emotional exhaustion, this would indicate an opportunity to prioritize reducing frustration through better design, training, and implementation as a mechanism to combat burnout [27].

We sought to quantify health care workers' frustration with technology and its relationship with emotional exhaustion, after controlling for measures of work-life integration (WLI) that may indicate excessive job demands. We hypothesized that higher frustration with technology corresponds to higher emotional exhaustion.

Methods

Study Design and Population

This cross-sectional, observational study is a secondary analysis of the Safety, Communication, Operational Reliability, and Engagement (SCORE) survey distributed via email through the Michigan Health and Hospital Association Keystone Center in 2015 as part of their routine patient safety and quality measurement, allowing a single response per user [28,29]. All employees working 0.5 full-time equivalents or higher in any Michigan hospital for 4 consecutive weeks prior to the survey administration were invited to participate. Confidentiality was assured to the respondents, participation in the survey was voluntary, and no incentives were offered. No questions were randomized or adapted to responses in real-time, no completeness check was enforced, and reviewing of answers was allowed. No cookies, internet protocol address checks, or log files were used to exclude responses. All surveys that contained responses to the scales measuring WLI and emotional exhaustion (described below) were analyzed. Surveys with "not applicable" or missing responses to either of these two scales were excluded. This study was not considered human subjects research by Stanford University and was approved by the institutional review board at Duke University Medical Center (Pro00033155).

Measures

The SCORE survey measures common workplace issues and work setting norms [28,29], including WLI, Emotional Exhaustion, Local Leadership, Learning Environment, Burnout Climate, Teamwork Climate, and Safety Climate scales [28,30,31]. SCORE also contains workforce engagement subscales and demographic questions (ie, number of years in specialty, job position, shift type, and length).

WLI and Frustration With Technology

We assessed WLI using a scale primarily focused on tangible frequencies of activities reflecting the interaction between work and personal responsibilities [28,29,32,33]. Each WLI item begins with the phrase "During the past week, how often did this occur?" The WLI items are as follows: (1) skipped a meal, (2) ate a poorly balanced meal, (3) worked through a shift with no breaks, (4) arrived home late from work, (5) had difficulty sleeping, (6) slept less than 5 hours in a night, and (7) changed personal/family plans due to work.

This WLI scale was originally validated as a 7-item scale as described above (Cronbach α=.79 in a validation study [29] and α=.81 in the current data set) and later updated with an additional eighth item assessing the frequency that one "felt frustrated with technology" as a key indicator of the ability of technology to facilitate efficient workflows and minimize conflicts between work and personal responsibilities (Cronbach α=.83 in a validation study [28] and α=.81 in the current data set). The full 8-item WLI scale is the current standard, but for

XSL•FO
RenderX

the purposes of this study, we calculated WLI scores using the previously validated 7-item WLI scale, separately considering the additional item relating to frustration with technology as our primary independent variable of interest.

Each item spans a 4-point Likert scale ("rarely or none of the time," "some or a little of the time," "occasionally or a moderate amount of time," and "all of the time"). For ease of interpretation, when assessing global correlations, we transposed the mean score of the 7 WLI items onto a scale of 0 to 100 and reversed the valence, with 100 indicating a favorable score (high WLI) and 0 indicating a poor score (low WLI). We similarly transposed frustration with technology onto a scale of 0 to 100, with 100 indicating a poor score (high frustration) and 0 indicating a favorable score (low frustration). For a secondary analysis comparing aggregated scores by work setting, we divided work settings into four quartiles based on the mean frustration with technology score.

Emotional Exhaustion

The 5-item Emotional Exhaustion scale (Cronbach α=.92) of the SCORE survey is composed of 4 items adapted from the Emotional Exhaustion subscale of the Maslach Burnout Inventory (r=0.96-0.98 with the 9-item scale) and a fifth item developed to align with the job demands-resources model of burnout [34-36]. This scale has been validated for use among health care workers as a burnout metric (eg, "Events in this work setting affect my life in an emotionally unhealthy way"), with demonstrated content, internal consistency, and consequence validity for this purpose [28-31,37]. Responses span a 5-point Likert scale from "disagree strongly" to "agree strongly." We calculated each individual's emotional exhaustion score by transposing the mean score of the five items onto a scale of 0 to 100, in line with a previous study [38].

The complete SCORE survey alongside derivation, scoring procedures, and reliability data for each of its scales is available on the internet [39].

Classifications

Individual responses were also categorized by their specific work setting, based on self-reported work location, such as St. Elsewhere Hospital 5 South, Pleasantview Pediatrics Clinic, and Mercy Health Systems Billing Department (fictional names). Each work setting thus reflects a grouping of respondents who work together as a team, regardless of each respondent's individual role. Work settings were classified as direct patient care (clinical) or indirect patient care (nonclinical, including administrative or billing departments). Work settings providing direct patient care were further classified as either intensive/emergency or acute; surgical or medical; and inpatient, outpatient, or mixed inpatient/outpatient. To maintain confidentiality and reduce risk of response bias from small samples, respondents from work settings with fewer than 5 total respondents were excluded from correlation and regression analyses.

Statistical Analysis

Descriptive statistics are presented as mean and SD values, or frequencies and percentages as appropriate. We compared group means by performing two-tailed t tests. We evaluated agreement in frustration with technology scores within work settings by using weighted Cohen kappa agreement analysis. As a first step insight-generation analysis, we evaluated Pearson correlations among survey items by using mean scores aggregated by work setting and weighted by the number of responses, avoiding the assumptions of identically distributed observations across work settings and of nested results (eg, health care workers within work settings) [40]. In our primary analysis, we evaluated the independent relations between frustration with technology score, other measures of WLI, and the outcome of emotional exhaustion by using a single mixed-effect generalized linear regression model, with work setting as random intercept, and job position, number of years in specialty, and work setting classifiers as fixed effects. We also performed, as a sensitivity analysis, a secondary validation to control for any available potential confounding factor that the primary regression may have omitted. We leveraged the statistical machine learning method lasso [41,42] to select relevant covariates from a large set of 36 potential covariates and re-ran our regression. Analyses were performed using Stata/IC software (version 15.1; StataCorp LLC). We used simple Bonferroni correction to account for multiple comparisons. With a total of 11 comparisons (8 independent items in the regression model, plus 3 t tests of adjacent quartiles) and a desired family-wise error rate of <0.05, two-tailed P values <.0045 were considered statistically significant.

Results

Of the 23,853 distributed surveys, 16,797 were returned (70.4% response rate). Of these, 915 indicated that technology use was "not applicable" to them and 377 had incomplete responses, resulting in 15,505 complete responses for further analysis. Descriptive statistics are presented in Table 1. The most frequently represented positions among all respondents (N=15,505) were nurses (n=4316, 27.8%), technologists and technicians (n=1890, 12.2%), and administrative support personnel (n=1800, 11.6%). The majority of respondents (n=10,284, 66.3%) reported 5 or more years in their current specialty. Nearly half (n=7286, 47.0%) of the respondents were from units not providing direct patient care, 9.0% (n=1398) were from units providing intensive or emergency care, and 10.0% (n=1559) were from units providing surgical care. Of the 1140 work settings represented, 818 (71.8 %) had 5 or more unique respondents and were included in regression analyses.

XSL•FO
RenderX

Table 1. Characteristics of survey respondents (N=15,505) from 1140 different work settings (818 work settings had 5 or more respondents).

Characteristic	Participant, n (%)
Position	
Nurse	4316 (27.8)
Technologist/Technician	1890 (12.2)
Admin support	1800 (11.6)
Admin/Manager	1238 (8.0)
Clinical support	839 (5.4)
Therapist	696 (4.5)
Nurses' aide	626 (4.0)
Physician	431 (2.9)
Environmental support	288 (1.9)
Pharmacist	226 (1.5)
Physician assistant	105 (0.7)
Other	3050 (19.7)
Years in specialty	
0-2	3056 (20.0)
3-4	1933 (12.7)
5-10	3374 (22.1)
11-20	3684 (24.1)
21 or more	3226 (21.1)
Setting	
Indirect patient care	7286 (47.0)
Direct patient care	8219 (53.0)
Acute care	6821 (44.0)
ICU[a]	1398 (9.0)
Medical	6660 (43.0)
Surgical	1559 (10.0)
Inpatient	4344 (28.0)
Mixed	3045 (19.6)
Outpatient	830 (5.4)
Shift	
Day	10,979 (70.8)
Night	2250 (14.5)
Swing	817 (5.3)
Other	1214 (7.8)
Shift length (hours)	
8	7889 (50.9)
10	1272 (8.2)
12	4091 (26.4)
Flexible	874 (5.6)
Other	1211 (7.8)
Frustration with technology	
Rarely	6310 (40.7)

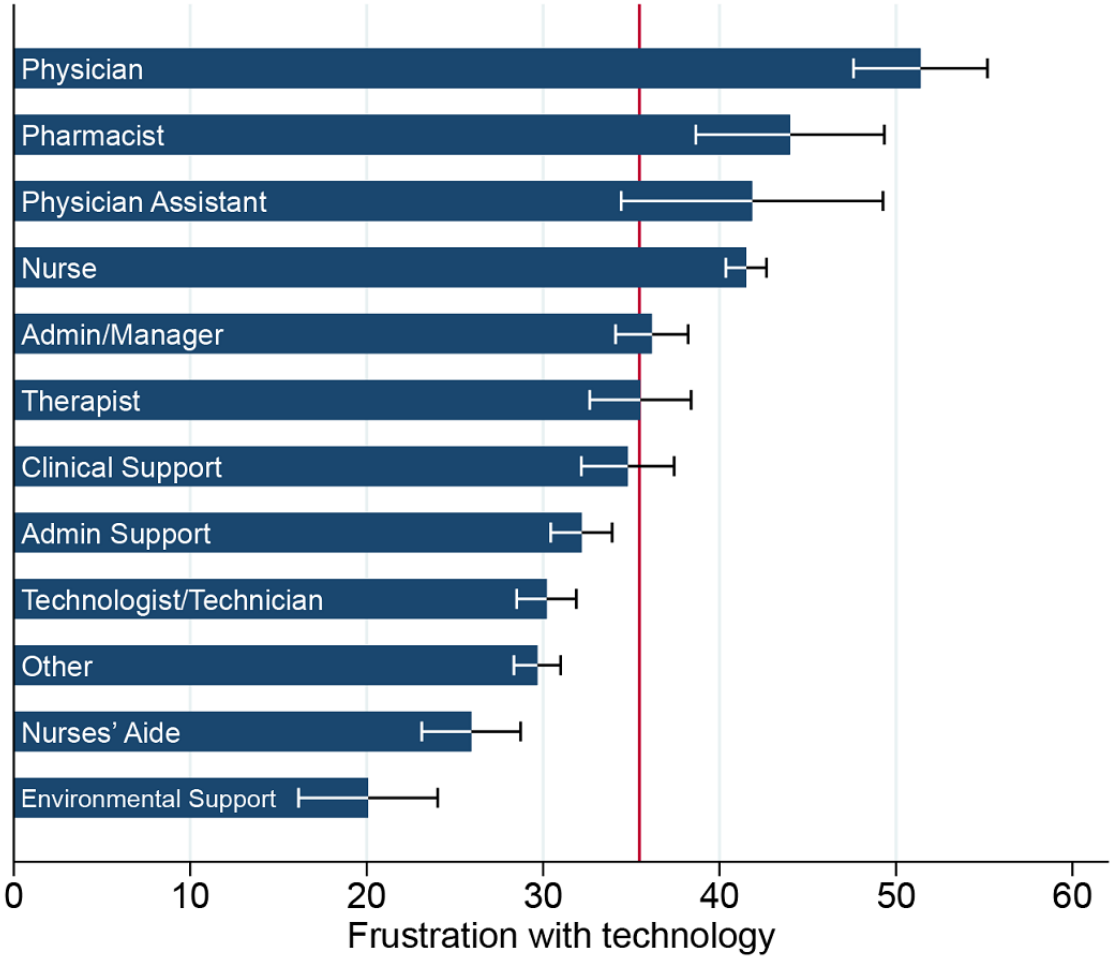
XSL•FO
RenderX

Characteristic	Participant, n (%)
A little	4130 (26.6)
Occasionally	2815 (18.2)
Always	2250 (14.5)

[a]ICU: intensive care unit.

WLI scores ranged from 0 to 100 (higher score favorable), with a mean score of 68.4 (SD 23.4) and median score of 71.4 (IQR 52.4-85.7). Separately, frustration with technology was reported as “none/rarely” by 6310 (40.7%) of the 15,505 respondents, “some/a little” by 4130 (26.6%), “occasionally/moderate” by 2815 (18.2%), and “all the time” by 2250 (14.5%) respondents. Frustration with technology scores ranged from 0 to 100 (lower score favorable), with a mean score of 35.0 (SD 35.9), and the score was higher among direct clinical care providers (mean 36.8, SD 36.4) than indirect providers (mean 32.9, SD 35.3; P<.001). The distribution of frustration with technology scores among 818 work settings, with corresponding WLI scores, is shown in Multimedia Appendix 1. The mean frustration with technology scores by job type are presented in Figure 1; the highest scores were reported by physicians, pharmacists, physician assistants, and nurses. Agreement in frustration with technology within work settings was low, with a combined weighted Cohen κ=0.04. Emotional exhaustion scores ranged from 0 to 100 (lower score favorable), with a mean score of 41.6 (SD 30.7).

Figure 1. Frustration with technology scores by job position. Data shown as mean values and 95% confidence limits of the mean, with the reference line at a population mean of 35.03.



Work setting correlations among frustration with technology, WLI items, and emotional exhaustion responses are illustrated in Multimedia Appendix 2 and tabulated in Multimedia Appendix 3. Frustration with technology was positively correlated with the full Emotional Exhaustion scale (r=0.35) as well as each individual item on the scale (r=0.29-0.36). The reverse-transposed WLI scale was negatively correlated with the Emotional Exhaustion scale and its individual items (r=–0.55 to –0.63). Each individual WLI item was correlated with the Emotional Exhaustion scale and its individual items, with the smallest correlations for “arrived home late from work”

XSL•FO
RenderX

(*r*=0.32-0.41) and the largest correlations for "had difficulty sleeping" (*r*=0.57-0.65).

In our primary analysis, frustration with technology and six of the seven WLI items were independently related to emotional exhaustion in multivariable modeling, each associating with a 0.34- to 2.06-point increase in the emotional exhaustion score for each 10-point change (on a 100-point scale), as shown in Table 2. Frustration with technology was associated with a 1.23-point (95% CI 1.07-1.38) increase in the emotional exhaustion score with each 10-point change, and it was second only to *difficulty sleeping* out of the WLI items most strongly associated with emotional exhaustion. For example, an increase in frustration with technology score from 30 to 40 would correspond to a 1.23-point increase in the emotional exhaustion score, all else being equal. The frequency of sleeping less than 5 hours a night was the only WLI item that was not independently associated with emotional exhaustion. Results were similar when stratified by direct patient care versus indirect patient care. The results of our sensitivity analysis, as shown in Multimedia Appendix 4, are aligned with our primary regression model.

Table 2. Frustration with technology and work-life integration as independent predictors of emotional exhaustion.

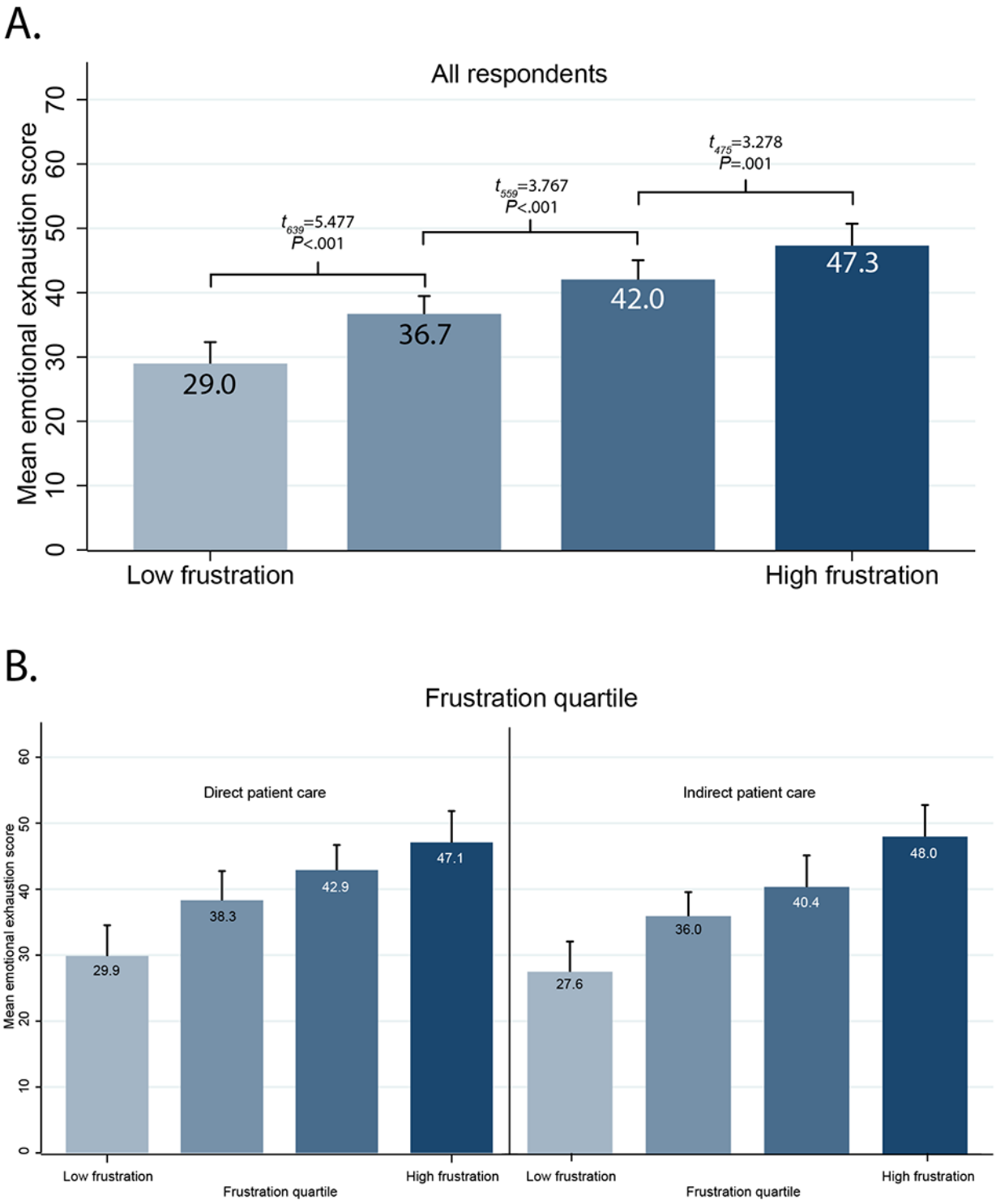
Work-Life Integration scale item	β[a]	95% CI	*P* value
Felt frustrated by technology	1.23	1.07 to 1.38	<.001
During the past week, how often did this occur?			
Had difficulty sleeping	2.06	1.88 to 2.25	<.001
Changed personal/family plans because of work	.99	0.80 to 1.18	<.001
Worked through a day/shift without any breaks	.87	0.69 to 1.05	<.001
Arrived home late from work	.82	0.64 to 1.00	<.001
Ate a poorly balanced meal	.67	0.49 to 0.85	<.001
Skipped a meal	.34	0.14 to 0.54	.001
Slept less than 5 hours in a night	.01	–0.18 to 0.19	.94

[a]Estimates via a single multivariable mixed model with work setting as random intercept. Beta coefficients reflect the change in emotional exhaustion score for each 10-point increase in frustration or work-life integration item (100-point scale) evaluated among 12,528 respondents in 818 work settings, adjusted for job type, years of experience, patient care type (intensive care vs not, surgical vs not, inpatient vs not), and direct patient care vs indirect patient care.

Our secondary analysis relating frustration with technology and emotional exhaustion aggregated within work settings is shown in Figure 2. Work settings with higher mean frustration with technology scores (distributed into quartiles) had higher emotional exhaustion scores. Mean emotional exhaustion scores ranged from 29.0 (SD 18.7, 95% CI 25.6-32.3) in the lowest quartile to 47.3 (SD 19.1, 95% CI 43.9-50.7) in the highest quartile. Results were similar when stratified by direct patient care versus indirect patient care.

XSL•FO
RenderX

Figure 2. Emotional exhaustion scores, stratified by quartile of the technology frustration scores for each work setting, shown for all respondents (A) and stratified by direct patient care versus indirect patient care (B). Data are shown as mean values and upper 95% confidence limits, with results of *t* tests of adjacent quartiles.



Discussion

This study found that frustration with technology varies with health care worker role and among individuals within work settings. Frustration with technology and 6 of 7 WLI items are independently associated with emotional exhaustion. Although frustration with technology was higher among direct clinical providers, similar relationships with emotional exhaustion were apparent for respondents engaged in direct patient care compared to those engaged in indirect patient care.

These results highlight and build on the evidence relating health care workers' user experience with well-being outcomes [16,43,44]. HIT differs from consumer technology in that the purchaser (eg, health care administration) often has different

XSL•FO
RenderX

priorities than the end users (health care workers), making user experience less incentivized as a driver in technology development. Although the purchaser's incentives (eg, improving patient care quality and communication, facilitating accurate reimbursement, and developing research and analytics infrastructure) are undeniably important, our findings highlight the need for attention to user experience with HIT [45]. Recently, physician-reported EHR usability was reported to be in the bottom 10th percentile relative to system usability scores in other industries, but physicians reporting better EHR usability had lower odds of burnout [26].

Frustration with technology may arise from discrepancies between expectations and reality [24]. Particularly in settings in which EHRs or other HITs were rapidly implemented, health care workers' expectations of the benefits of these technologies may not have been accurately set, features may not have been fully explained, or efficient use of the interface may not have been taught [46,47]. The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 that incentivized the adoption of HIT was effective at encouraging transition to EHRs, but rapid adoption may have precluded adequate attention to end-user input, development of interoperability standards, setting realistic expectations, or education on the use of new technologies [1,48-50].

Emotional exhaustion, the key construct of burnout evaluated in this study, reflects a depleted state arising from excessive demands, continuous stress, or insufficient resources [34,35,51-54]. If an EHR requires many clicks or otherwise inefficient workflows to accomplish tasks, this would translate to increased job demands. Alternatively, if inadequate training results in a health care worker not knowing how to use the interface effectively, this lack of knowledge translates to insufficient resources to use the available tools. In both cases, frustration with technology may develop as an indicator of imbalanced job demands and resources.

In our primary analysis, frustration with technology was found to be associated with burnout independently of other markers of WLI. This association was not as strong as the one observed for difficulty sleeping, consistent with observations that sleep disturbance may itself be an indicator of psychological distress [55]. However, it was similar in magnitude to the associations for other items reflecting excessive workload or workplace inefficiencies, such as changing personal/family plans because of work, working a shift without any breaks, or arriving home late from work. This pattern of findings suggests that frustration with technology is more closely related to an imbalance between job demands and resources, and it is less similar to psychological distress markers.

At the work-setting level, average frustration with technology was significantly associated with higher emotional exhaustion, which indicates a potential climate-like effect of frustration with technology, similar to that observed for WLI [28]. However, agreement among respondents from the same work setting was low, suggesting that much of frustration with technology is rooted at the individual level rather than the work-setting level (ie, the individual's experience with technology rather than problems with the technology itself). Even work settings with the highest average frustration with technology scores still contained individuals reporting low frustration, suggesting against the notion that frustration is inevitable or inextricably linked to HIT. Different individuals may have vastly different experiences using the same technology, related to their specific worker tasks, training, personal expectations, comfort with the technology, and acceptance of change [56].

Of note, frustration with technology was only moderately correlated with frustration with one's job in general, indicating that frustration with technology can occur without generalized job frustration and that individuals may remain motivated to take action to reduce their frustrations. This pattern of findings suggests that ensuring all individuals receive adequate support and training may be effective at reducing frustration with technology, rather than focusing only on making extensive changes to the technology or the user interface itself.

Several practical steps may be taken to reduce the frequency of frustration with technology among health care workers. Although details will require tailoring to specific settings, tasks, and technologies available, our findings suggest an approach of educating and supporting individuals experiencing the most frequent frustrations. Recent studies have described the use of supplemental EHR training to improve comfort with technology and ability to work more efficiently [57,58]. Interventions at the work-setting level may also include measuring and reducing individual workloads, such as employing scribes to assist with clinical documentation, transitioning to team-based documentation and inbox management, or automating data-entry tasks [59-72]. Finally, interventions to reduce workloads placed on the system could have the broadest benefits, but it would likely require changes to current payment structures that promote lengthy documentation and labor-intensive payment authorizations [73].

This study must be interpreted in the context of its design. As a cross-sectional, observational study, it cannot determine causality or directionality of the observed correlations (ie, whether frustration with technology induces burnout, burnout amplifies frustrations, and/or an external factor influences both). Although we were able to adjust for many potential confounders in our regression model, it is possible that residual confounding from unobserved variables such as age or prior experience with technology remains. Emotional exhaustion was also evaluated, but it does not capture other manifestations of burnout such as depersonalization. The 100-point emotional exhaustion scale we used differs from scores generated by other burnout instruments and cannot be used to directly compare effect sizes from studies using different instruments; however, methods to approximate estimates from disparate instruments have been previously described [74]. Although we used the 7-item WLI scale in this study, the 8-item version is commonly used, internally consistent (Cronbach α=.81 for this sample), and appears to add a unique element to the WLI assessment as evidenced by the results presented here.

The response rate of over 70% compares favorably with other studies of this magnitude and exceeds commonly accepted thresholds for survey-based research; however, there may remain some sampling bias, and it is possible that physicians in

XSL•FO
RenderX

particular were relatively underrepresented in this sample [75,76]. Similarly, although the survey was confidential, as with any self-reported measures, the responses may be susceptible to recall bias or social desirability bias. The survey respondents reflect a wide variety of health care roles and work settings, with varying interactions with technology. Furthermore, even though many of the direct and indirect patient care roles included in this study heavily feature EHR use, these and other respondents may have referenced other use of technology in their responses, and we did not conduct interviews to further characterize their responses. Prior qualitative research has found that specific drivers of HIT frustration vary among individuals, tasks, and settings [47,77]. Our findings are thus reflective of the overall conceptual relationship between frustration and emotional exhaustion but are unable to provide conclusions regarding any particular piece of technology or source of frustration. Additional research will be necessary to further delineate the specific sources and scope of frustration with technology across health care worker roles, as these data may provide more granular insights of potential interventions to reduce frustrations. Although this survey was administered within a single US state, the Michigan Health and Hospital Association Keystone Center includes all 175 hospitals from 20 health systems within Michigan, making our results likely to be generalizable to community and academic hospitals across the United States.

It remains unknown whether reducing frustrations with technology through improved training, updated interfaces, or redistributed tasks will be effective in reducing burnout. Longitudinal observational studies may enhance our understanding of the directionality of these relationships, but prospective trials will be needed to fully evaluate the effect of interventions to improve health care worker user experience and well-being.

In conclusion, frustration with technology and difficulty sleeping were the biggest WLI factors associated with emotional exhaustion across direct and indirect patient care settings. Interventions designed to reduce health care workers' frustration with technology and improve other aspects of WLI may be effective strategies to reduce burnout among health care workers.

Acknowledgments

This study was supported by the Agency for Healthcare Research and Quality (grant K08 HS027837, awarded to DT), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant R01 HD084679-01, awarded to JBS and JP), and the Stanford Maternal and Child Health Research Institute (DT). The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

TDS is co-inventor of the Well-Being Index instruments and the Participatory Management Leadership Index. Mayo Clinic holds the copyright to these instruments and has licensed them for use outside Mayo Clinic. Mayo Clinic pays TDS a share of royalties received. As an expert in clinician well-being, TDS frequently presents grand round and keynote presentations and advises medical centers on clinician well-being. He receives honorarium for some of these engagements.

Multimedia Appendix 1
Frustration with technology and work-life integration scores for 818 individual work settings, ordered by mean frustration with technology scores. Lower panel shows mean work-life integration scores for the corresponding units, suggesting a negative unadjusted correlation. Data are shown as mean values and 95% confidence limits.
[DOCX File , 319 KB - jmir_v23i7e26817_app1.docx]

Multimedia Appendix 2
Work-setting frustration with technology scores in relation to emotional exhaustion scores. For ease of interpretation, values are shown as means and 50% confidence limits for each work setting (total of 818 work settings with 5 or more respondents). CLM: confidence limits of the means.
[DOCX File , 215 KB - jmir_v23i7e26817_app2.docx]

Multimedia Appendix 3
Item-level correlations with emotional exhaustion items for 818 work settings.
[DOCX File , 14 KB - jmir_v23i7e26817_app3.docx]

Multimedia Appendix 4
Sensitivity analysis. Multivariable regression with independent variables chosen using lasso.
[DOCX File , 14 KB - jmir_v23i7e26817_app4.docx]

References

XSL•FO
RenderX

1. Washington V, DeSalvo K, Mostashari F, Blumenthal D. The HITECH era and the path forward. N Engl J Med 2017 Sep 07;377(10):904-906. [doi: 10.1056/NEJMp1703370] [Medline: 28877013]
2. Martin SA, Sinsky CA. The map is not the territory: medical records and 21st century practice. Lancet 2016 Oct 22;388(10055):2053-2056. [doi: 10.1016/S0140-6736(16)00338-X] [Medline: 27125861]
3. Berenson RA, Basch P, Sussex A. N Engl J Med 2011 May 19;364(20):1892-1895. [doi: 10.1056/NEJMp1102099] [Medline: 21591942]
4. Halamka JD, Tripathi M. The HITECH era in retrospect. N Engl J Med 2017 Sep 07;377(10):907-909. [doi: 10.1056/NEJMp1709851] [Medline: 28877012]
5. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
6. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
7. Spinelli WM, Fernstrom KM, Britt H, Pratt R. "Seeing the patient is the joy:" a focus group analysis of burnout in outpatient providers. Fam Med 2016 Apr;48(4):273-278. [Medline: 27057605]
8. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 06;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
9. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Aff (Millwood) 2017 Apr 01;36(4):655-662 [FREE Full text] [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
10. Shanafelt TD, West CP, Sinsky C, Trockel M, Tutty M, Satele DV, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general US working population between 2011 and 2017. Mayo Clin Proc 2019 Sep;94(9):1681-1694 [FREE Full text] [doi: 10.1016/j.mayocp.2018.10.023] [Medline: 30803733]
11. Han S, Shanafelt TD, Sinsky CA, Awad KM, Dyrbye LN, Fiscus LC, et al. Estimating the attributable cost of physician burnout in the United States. Ann Intern Med 2019 Jun 04;170(11):784-790. [doi: 10.7326/M18-1422] [Medline: 31132791]
12. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. JAMA Intern Med 2017 Dec 01;177(12):1826-1832. [doi: 10.1001/jamainternmed.2017.4340] [Medline: 28973070]
13. Tran B, Lenhart A, Ross R, Dorr DA. Burnout and EHR use among academic primary care physicians with varied clinical workloads. AMIA Jt Summits Transl Sci Proc 2019;2019:136-144 [FREE Full text] [Medline: 31258965]
14. Robertson SL, Robinson MD, Reid A. Electronic health record effects on work-life balance and burnout within the I Population Collaborative. J Grad Med Educ 2017 Aug;9(4):479-484 [FREE Full text] [doi: 10.4300/JGME-D-16-00123.1] [Medline: 28824762]
15. How doctors feel about electronic health records. National Physician Poll by The Harris Poll.: Stanford Medicine; 2018. URL: https://med.stanford.edu/ehr/electronic-health-records-poll-results.html [accessed 2021-06-15]
16. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
17. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc 2014 Feb;21(e1):e100-e106 [FREE Full text] [doi: 10.1136/amiajnl-2013-001875] [Medline: 24005796]
18. Tawfik DS, Phibbs CS, Sexton JB, Kan P, Sharek PJ, Nisbet CC, et al. Factors associated with provider burnout in the NICU. Pediatrics 2017 May;139(5):e20164134 [FREE Full text] [doi: 10.1542/peds.2016-4134] [Medline: 28557756]
19. Tai-Seale M, Dillon EC, Yang Y, Nordgren R, Steinberg RL, Nauenberg T, et al. Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. Health Aff (Millwood) 2019 Jul;38(7):1073-1078. [doi: 10.1377/hlthaff.2018.05509] [Medline: 31260371]
20. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. J Am Med Inform Assoc 2020 Apr 01;27(4):531-538 [FREE Full text] [doi: 10.1093/jamia/ocz220] [Medline: 32016375]
21. Hilliard R, Haskell J, Gardner R. Are specific elements of electronic health record use associated with clinician burnout more than others? J Am Med Inform Assoc 2020 Jul 01;27(9):1401-1410. [doi: 10.1093/jamia/ocaa092] [Medline: 32719859]
22. National Academy of Medicine. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. In: National Academies of Sciences. Washington, DC: The National Academies Press; 2019.
23. Lawson R. Frustration: the development of a scientific concept. United Kingdom: Macmillan; 1965.
24. Opoku-Boateng GA. User frustration in HIT interfaces: exploring past Hci research for a better understanding of clinicians' experiences. AMIA Annu Symp Proc 2015;2015:1008-1017 [FREE Full text] [Medline: 26958238]
25. Zayas-Cabán T, White PJ. The national health information technology human factors and ergonomics agenda. Appl Ergon 2020 Jul;86:103109. [doi: 10.1016/j.apergo.2020.103109] [Medline: 32342896]

XSL•FO
RenderX

26. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clin Proc 2020 Mar;95(3):476-487 [FREE Full text] [doi: 10.1016/j.mayocp.2019.09.024] [Medline: 31735343]
27. Friedberg M, Chen P, Van BK, Aunon F, Pham C, Caloyeras J. Factors affecting physician professional satisfaction and their implication for patient care, health systems, and health policy. Santa Monica, CA: RAND Corporation; 2013.
28. Schwartz SP, Adair KC, Bae J, Rehder KJ, Shanafelt TD, Profit J, et al. Work-life balance behaviours cluster in work settings and relate to burnout and safety culture: a cross-sectional survey analysis. BMJ Qual Saf 2019 Feb;28(2):142-150 [FREE Full text] [doi: 10.1136/bmjqs-2018-007933] [Medline: 30309912]
29. Sexton JB, Schwartz SP, Chadwick WA, Rehder KJ, Bae J, Bokovoy J, et al. The associations between work-life balance behaviours, teamwork climate and safety climate: cross-sectional survey introducing the work-life climate scale, psychometric properties, benchmarking data and future directions. BMJ Qual Saf 2017 Aug;26(8):632-640 [FREE Full text] [doi: 10.1136/bmjqs-2016-006032] [Medline: 28008006]
30. Sexton JB, Helmreich RL, Neilands TB, Rowan K, Vella K, Boyden J, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. BMC Health Serv Res 2006 Apr 03;6:44 [FREE Full text] [doi: 10.1186/1472-6963-6-44] [Medline: 16584553]
31. Profit J, Sharek PJ, Amspoker AB, Kowalkowski MA, Nisbet CC, Thomas EJ, et al. Burnout in the NICU setting and its relation to safety culture. BMJ Qual Saf 2014 Oct;23(10):806-813 [FREE Full text] [doi: 10.1136/bmjqs-2014-002831] [Medline: 24742780]
32. Adair KC, Quow K, Frankel A, Mosca PJ, Profit J, Hadley A, et al. The Improvement Readiness scale of the SCORE survey: a metric to assess capacity for quality improvement in healthcare. BMC Health Serv Res 2018 Dec 17;18(1):975 [FREE Full text] [doi: 10.1186/s12913-018-3743-0] [Medline: 30558593]
33. Tawfik DS, Shanafelt TD, Dyrbye LN, Sinsky CA, West CP, Davis AS, et al. Personal and professional factors associated with work-life integration among US physicians. JAMA Netw Open 2021 May 03;4(5):e2111575 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.11575] [Medline: 34042994]
34. Bakker AB, Demerouti E. Job demands-resources theory: taking stock and looking forward. J Occup Health Psychol 2017 Jul;22(3):273-285. [doi: 10.1037/ocp0000056] [Medline: 27732008]
35. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001 Jun;86(3):499-512. [Medline: 11419809]
36. Block M, Ehrenworth JF, Cuce VM, Ng'ang'a N, Weinbach J, Saber S, et al. Measuring handoff quality in labor and delivery: development, validation, and application of the Coordination of Handoff Effectiveness Questionnaire (CHEQ). Jt Comm J Qual Patient Saf 2013 May;39(5):213-220. [doi: 10.1016/s1553-7250(13)39028-x] [Medline: 23745480]
37. Sexton JB, Adair KC, Leonard MW, Frankel TC, Proulx J, Watson SR, et al. Providing feedback following Leadership WalkRounds is associated with better patient safety culture, higher employee engagement and lower burnout. BMJ Qual Saf 2018 Dec;27(4):261-270 [FREE Full text] [doi: 10.1136/bmjqs-2016-006399] [Medline: 28993441]
38. Sexton JB, Adair KC. Forty-five good things: a prospective pilot study of the Three Good Things well-being intervention in the USA for healthcare worker emotional exhaustion, depression, work-life balance and happiness. BMJ Open 2019 Mar 20;9(3):e022695 [FREE Full text] [doi: 10.1136/bmjopen-2018-022695] [Medline: 30898795]
39. Sexton J. SCORE: Assessment of your work setting Safety, Communication, Operational Reliability, and Engagement. Technical Report 16-8. 2017. URL: https://www.hsq.dukehealth.org/files/2019/05/SCORE_Technical_Report_5.14.19.pdf [accessed 2021-06-16]
40. Dowding I, Haufe S. Powerful statistical inference for nested data using sufficient summary statistics. Front Hum Neurosci 2018;12:103 [FREE Full text] [doi: 10.3389/fnhum.2018.00103] [Medline: 29615885]
41. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 2018 Dec 05;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
42. Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. SIAM Rev 2001 Jan;43(1):129-159. [doi: 10.1137/s003614450037906x]
43. Harris DA, Haskell J, Cooper E, Crouse N, Gardner R. Estimating the association between burnout and electronic health record-related stress among advanced practice registered nurses. Appl Nurs Res 2018 Oct;43:36-41. [doi: 10.1016/j.apnr.2018.06.014] [Medline: 30220361]
44. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 01;26(2):106-114. [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
45. American Medical Association. Improving care: Priorities to improve electronic health record usability. Chicago, IL: American Medical Association; 2014.
46. Jamoom EW, Heisey-Grove D, Yang N, Scanlon P. Physician opinions about EHR use by EHR experience and by whether the practice had optimized its EHR use. J Health Med Inform 2016 Jul 30;7(4):1000240 [FREE Full text] [doi: 10.4172/2157-7420.1000240] [Medline: 27800279]

XSL•FO
RenderX

47. Magsamen-Conrad K, Checton M. Technology and health care: efficiency, frustration, and disconnect in the transition to electronic medical records. GSTF J Media Commun 2014 Feb;1(2):23-27 [FREE Full text] [doi: 10.5176/2335-6618_1.2.16] [Medline: 25729754]
48. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc 2016 Mar;23(2):375-379. [doi: 10.1093/jamia/ocv103] [Medline: 26228764]
49. Cantwell E, McDermott K. making technology talk: how interoperability can improve care, drive effidcincy, and reduce waste. Healthc Financ Manage 2016 May;70(5):70-76. [Medline: 27382711]
50. Pronovost PJ, Johns MME, Palmer S, Bono RC, Fridsma DB, Gettinger A, et al, editors. Procuring interoperability: Achieving high-quality, connected, and person-centered care. Washington, DC: National Academy of Sciences; 2018.
51. Maslach CJ, Leiter M. Maslach Burnout Inventory Manual. 3rd ed. Palo Alto, CA: Consulting Psychologists Press; 1996.
52. Profit J, Etchegaray J, Petersen LA, Sexton JB, Hysong SJ, Mei M, et al. The Safety Attitudes Questionnaire as a tool for benchmarking safety culture in the NICU. Arch Dis Child Fetal Neonatal Ed 2012 Mar;97(2):F127-F132 [FREE Full text] [doi: 10.1136/archdischild-2011-300612] [Medline: 22337935]
53. Lee RT, Ashforth BE. A meta-analytic examination of the correlates of the three dimensions of job burnout. J Appl Psychol 1996 Apr;81(2):123-133. [doi: 10.1037/0021-9010.81.2.123] [Medline: 8603909]
54. Bakker AB, Demerouti E, Euwema MC. Job resources buffer the impact of job demands on burnout. J Occup Health Psychol 2005 Apr;10(2):170-180. [doi: 10.1037/1076-8998.10.2.170] [Medline: 15826226]
55. Chueh K, Chen K, Lin Y. Psychological distress and sleep disturbance among female nurses: anxiety or depression? J Transcult Nurs 2021 Jan;32(1):14-20. [doi: 10.1177/1043659619881491] [Medline: 31625463]
56. Tutty MA, Carlasare LE, Lloyd S, Sinsky CA. The complex case of EHRs: examining the factors impacting the EHR user experience. J Am Med Inform Assoc 2019 Jul 01;26(7):673-677 [FREE Full text] [doi: 10.1093/jamia/ocz021] [Medline: 30938754]
57. DiAngi YT, Stevens LA, Halpern-Felsher B, Pageler NM, Lee TC. Electronic health record (EHR) training program identifies a new tool to quantify the EHR time burden and improves providers' perceived control over their workload in the EHR. JAMIA Open 2019 Jul;2(2):222-230 [FREE Full text] [doi: 10.1093/jamiaopen/ooz003] [Medline: 31984357]
58. Stevens LA, DiAngi YT, Schremp JD, Martorana MJ, Miller RE, Lee TC, et al. Designing an individualized ehr learning plan for providers. Appl Clin Inform 2017;8(3):924-935. [doi: 10.4338/ACI-2017-04-0054] [Medline: 30027541]
59. Contratto E, Romp K, Estrada CA, Agne A, Willett LL. Physician order entry clerical support improves physician satisfaction and productivity. South Med J 2017 Dec;110(5):363-368. [doi: 10.14423/SMJ.0000000000000645] [Medline: 28464179]
60. Guo U, Chen L, Mehta PH. Electronic health record innovations: helping physicians - one less click at a time. Health Inf Manag 2017 Sep;46(3):140-144. [doi: 10.1177/1833358316689481] [Medline: 28671038]
61. Bank AJ, Gage RM. Annual impact of scribes on physician productivity and revenue in a cardiology clinic. Clinicoecon Outcomes Res 2015;7:489-495 [FREE Full text] [doi: 10.2147/CEOR.S89329] [Medline: 26457055]
62. Gidwani R, Nguyen C, Kofoed A, Carragee C, Rydel T, Nelligan I, et al. Impact of scribes on physician satisfaction, patient satisfaction, and charting efficiency: a randomized controlled trial. Ann Fam Med 2017 Dec;15(5):427-433 [FREE Full text] [doi: 10.1370/afm.2122] [Medline: 28893812]
63. Heaton HA, Castaneda-Guarderas A, Trotter ER, Erwin PJ, Bellolio MF. Effect of scribes on patient throughput, revenue, and patient and provider satisfaction: a systematic review and meta-analysis. Am J Emerg Med 2016 Oct;34(10):2018-2028. [doi: 10.1016/j.ajem.2016.07.056] [Medline: 27534432]
64. Pozdnyakova A, Laiteerapong N, Volerman A, Feld LD, Wan W, Burnet DL, et al. Impact of medical scribes on physician and patient satisfaction in primary care. J Gen Intern Med 2018 Jul;33(7):1109-1115. [doi: 10.1007/s11606-018-4434-6] [Medline: 29700790]
65. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. Mayo Clin Proc 2016 Nov 18;92(1):129-146. [doi: 10.1016/j.mayocp.2016.10.004] [Medline: 27871627]
66. Linzer M, Poplau S, Babbott S, Collins T, Guzman-Corrales L, Menk J, et al. Worklife and wellness in academic general internal medicine: results from a national survey. J Gen Intern Med 2016 Sep;31(9):1004-1010 [FREE Full text] [doi: 10.1007/s11606-016-3720-4] [Medline: 27138425]
67. Sinsky CA, Willard-Grace R, Schutzbank AM, Sinsky TA, Margolius D, Bodenheimer T. In search of joy in practice: a report of 23 high-functioning primary care practices. Ann Fam Med 2013 May;11(3):272-278 [FREE Full text] [doi: 10.1370/afm.1531] [Medline: 23690328]
68. Gregory ME, Russo E, Singh H. Electronic health record alert-related workload as a predictor of burnout in primary care providers. Appl Clin Inform 2017 Jul 05;8(3):686-697 [FREE Full text] [doi: 10.4338/ACI-2017-01-RA-0003] [Medline: 28678892]
69. Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, et al. Metrics for assessing physician activity using electronic health record log data. J Am Med Inform Assoc 2020 Apr 01;27(4):639-643 [FREE Full text] [doi: 10.1093/jamia/ocz223] [Medline: 32027360]
70. Funk KA, Davis M. Enhancing the role of the nurse in primary care: the RN "co-visit" model. J Gen Intern Med 2015 Dec;30(12):1871-1873 [FREE Full text] [doi: 10.1007/s11606-015-3456-6] [Medline: 26194640]

XSL•FO
RenderX

71. Sinsky CA, Bodenheimer T. Powering-up primary care teams: advanced team care with in-room support. Ann Fam Med 2019 Jul 08;17(4):367-371. [doi: 10.1370/afm.2422]
72. Lyon C, English AF, Chabot Smith P. A team-based care model that improves job satisfaction. Fam Pract Manag 2018;25(2):6-11 [FREE Full text] [Medline: 29537246]
73. Mechanic D. Physician discontent: challenges and opportunities. JAMA 2003 Aug 20;290(7):941-946. [doi: 10.1001/jama.290.7.941] [Medline: 12928472]
74. Tawfik DS, Scheid A, Profit J, Shanafelt T, Trockel M, Adair KC, et al. Evidence relating health care provider burnout and quality of care: a systematic review and meta-analysis. Ann Intern Med 2019 Oct 15;171(8):555-567 [FREE Full text] [doi: 10.7326/M19-1152] [Medline: 31590181]
75. Gordon N, Davidoff F, Tarnow E, Reidenberg MM, Endriss K. A question of response rate. Science Editor 2002;25(1):25-26 [FREE Full text]
76. Hendra R, Hill A. Rethinking response rates: new evidence of little relationship between survey response rates and nonresponse bias. Eval Rev 2019 Oct;43(5):307-330. [doi: 10.1177/0193841X18807719] [Medline: 30580577]
77. De Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020 Aug 14;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]

Abbreviations

EHR: electronic health record
HIT: health information technology
HITECH: Health Information Technology for Economic and Clinical Health
SCORE: Safety, Communication, Operational Reliability, and Engagement
WLI: work-life integration

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by KC Wong, R Malka; comments to author 27.02.21; revised version received 08.04.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Tawfik DS, Sinha A, Bayati M, Adair KC, Shanafelt TD, Sexton JB, Profit J
Frustration With Technology and its Relation to Emotional Exhaustion Among Health Care Workers: Cross-sectional Observational Study
J Med Internet Res 2021;23(7):e26817
URL: https://www.jmir.org/2021/7/e26817
doi:10.2196/26817
PMID:34255674

©Daniel S Tawfik, Amrita Sinha, Mohsen Bayati, Kathryn C Adair, Tait D Shanafelt, J Bryan Sexton, Jochen Profit. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Efficacy and Cost-effectiveness of Promotion Methods to Recruit Participants to an Online Screening Registry for Alzheimer Disease Prevention Trials: Observational Study

Kenichiro Sato[1,2], MD, PhD; Yoshiki Niimi[3], MD, PhD; Ryoko Ihara[4], MD, PhD; Kazushi Suzuki[5], MD, PhD; Tatsushi Toda[1], MD, PhD; Atsushi Iwata[4], MD, PhD; Takeshi Iwatsubo[2], MD, PhD

[1]Department of Neurology, University of Tokyo Hospital, Tokyo, Japan

[2]Department of Neuropathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[3]Unit for Early and Exploratory Clinical Development, The University of Tokyo Hospital, Tokyo, Japan

[4]Department of Neurology, Tokyo Metropolitan Geriatric Medical Center Hospital, Tokyo, Japan

[5]Division of Neurology, Internal Medicine, National Defense Medical College, Saitama, Japan

Corresponding Author:
Takeshi Iwatsubo, MD, PhD
Department of Neuropathology
Graduate School of Medicine
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku
Tokyo,
Japan
Phone: 81 3 3815 5411
Email: iwatsubo@m.u-tokyo.ac.jp

Abstract

Background: Web-based screening may be suitable for identifying individuals with presymptomatic latent diseases for recruitment to clinical studies, as such people do not often visit hospitals in the presymptomatic stage. The promotion of such online screening studies is critical to their success, although it remains uncertain how the effectiveness of such promotion can differ, depending on the different promotion methods, domains of interest, or countries of implementation.

Objective: The Japanese Trial-Ready Cohort (J-TRC) web study is our ongoing online screening registry to identify individuals with presymptomatic Alzheimer disease (AD), aimed at facilitating the clinical trials for AD prevention. Within the first 9 months of its 2019 launch, the J-TRC web study recruited thousands of online participants via multiple methods of promotion, including press releases, newspaper advertisements, web advertisements, or direct email invitations. Here, we aimed to quantitatively evaluate efficacy and cost-effectiveness of each of these multimodal promotion methods.

Methods: We applied the vector-autoregression model to assess the degree of contribution of each type of promotion to the following target metrics: number of daily visitors to the J-TRC website, number of daily registrants to the J-TRC web study, daily rate of registration among visitors, daily rate of eligible participants among registrants, and median age of daily registrants. The average cost-effectiveness for each promotion method was also calculated using the total cost and the coefficients in the vector-autoregression model.

Results: During the first 9 months of the reviewed period from October 31, 2019 to June 17, 2020, there were 48,334 website visitors and 4429 registrations (9.16% of 48,334 visitors), of which 3081 (69.56%) were eligible registrations. Initial press release reports and newspaper advertisements had a marked effect on increasing the number of daily visitors and daily registrants. Web advertisements significantly contributed to the increase in daily visitors (P<.001) but not to the daily registrants, and it also lowered the rate of registrations and the median age of daily registrants. Website visitors from the direct email invitation sent to other cognitive registries seem to have registered with the highest reliability. The calculated average cost-effectiveness for the initial press release was US $24.60 per visitor and US $96.10 per registrant, while the calculated average cost-effectiveness for the newspaper advertisements was US $28.60 per visitor and US $227.90 per registrant.

Conclusions: Our multivariate time-series analysis showed that each promotion method had different features in their effect of recruiting participants to the J-TRC web study. Under the advertisement condition settings thus far, newspaper advertisements

XSL•FO
RenderX

and initial press releases were the most effective promotion methods, with fair cost-effectiveness that was equivalent to earlier online studies. These results can provide important suggestions for future promotions for the recruitment of presymptomatic participants to AD clinical trials in Japan.

(J Med Internet Res 2021;23(7):e26284) doi:10.2196/26284

KEYWORDS

online clinical study; promotion; advertisement; cost-effectiveness; Trial-Ready Cohort; preclinical Alzheimer disease; clinical trial; Alzheimer; dementia; recruitment

Introduction

Web-based clinical studies are rapidly growing in importance to become one of the promising methods of clinical research [1], especially, in the era of COVID-19 [2]. In addition to targeting already developed diseases, a web-based approach such as an online survey may also be particularly suitable to target presymptomatic, latent diseases [3-5] or for presymptomatic disease screening [6], because asymptomatic individuals cannot always be expected to visit hospitals of their own volition.

Study advertising to reach the target population is critical to the success of web-based clinical studies [1,7]. Currently, there are several types of methods available to promote clinical studies [3-6,8-13], such as advertisements in standard media like newspapers [14] or television [8], online advertisements (eg, Google, Yahoo, or Facebook) [1], social media posts (eg, Facebook, Twitter, or Instagram) [8], direct phone calls or emails [8,15], or word of mouth [9]. Which of these modalities may be more effective for the target population might vary depending on the disease domain of interest, design of the clinical study, or regional environments associated with the clinical studies [5]. For example, it is reported that Japanese people in their 60s regard the newspaper as the more reliable news media than the internet or television [16], which suggests that newspaper advertisements may be more suitable than online advertisements to reach older people and to appeal them to participate in clinical studies of age-related diseases. However, the current evidence on methods that recruit web-based clinical studies are mainly based on those associated with smoking cessation [8], mental health [1], hypertension [17], or cancer [18]. Evidence was especially limited in terms of their employment in the field of neurological diseases, except for a very few earlier studies using online recruitment for prevention trials of Alzheimer disease (AD) [19].

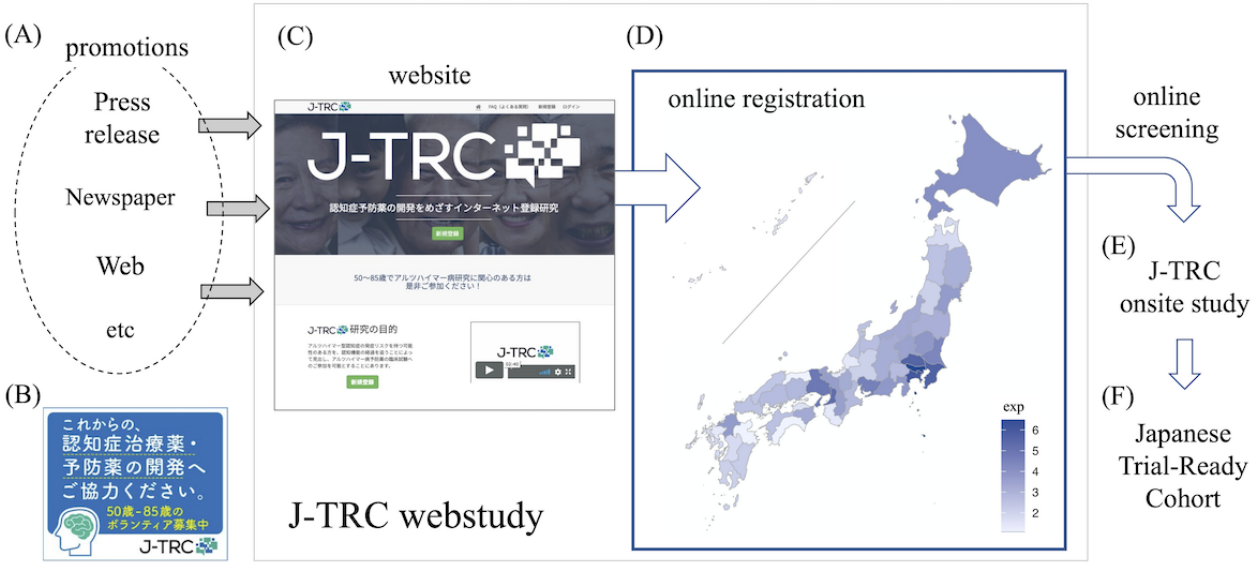
Individuals with preclinical AD, which corresponds to the presymptomatic stage of AD where patients exhibit the earliest pathological changes in the brain but without significant cognitive decline [20-23], have now been focused on as primary targets of AD clinical trials. As they are asymptomatic, AD researchers often recruit participants via an online screening registry. This method has a problem similar to earlier web-based clinical interventions in other domains: how to promote recruitment for preclinical AD screening studies with limited research funds while incorporating the characteristics of preclinical AD individuals as the target group.

Since 2019, we have been conducting the Japanese Trial-Ready Cohort (J-TRC) web study [24] (limited to Japanese domestic access only) (Figure 1), as one such online screening registry of preclinical AD [25]. Within the first 9 months of its 2019 launch, the J-TRC web study has recruited more than 4000 online participants from all over Japan [26] via multiple modalities of promotions (Figure 1A), including press releases, newspaper advertisements, web advertisements (Figure 1B), and direct email invitations sent to another related registry. Among the website visitors, those who were willing to participate in the study proceeded to study registration (Figure 1C) and were screened for their web-based cognitive performance every 3 months to further refer them to the successive in-person study (Figure 1E), aiming to build a large Japanese cohort of preclinical individuals ready for the AD clinical trials (Figure 1F).

Studies of the detailed features or differences in the effectiveness of these promotion methods in the context of the J-TRC web study promotion will be informative for the future planning of promotions for similar settings of presymptomatic online screening registries, regardless of the disease domain. However, in the J-TRC web study promotion, these promotion methods were conducted using a combined approach, with unignorable overlap in their timings during the same period. Thus, a direct comparison of the effectiveness between these different methods is impossible. Therefore, by applying a multivariate time-series analysis, we quantitatively evaluated the efficacy and cost-effectiveness of each of these promotion methods. Our study will also be informative to fill the gap in evidence about the difference in the effectiveness of several methods to promote web-based clinical studies between different countries (primarily between Japan and Western countries), since the earlier online clinical studies have come from Western countries where clinical trial environments are different from those in Japan.

XSL•FO
RenderX

Figure 1. Schematic outline of the study. Multiple methods (A) including web banner advertisement (B) were used to promote visitors and registrants to J-TRC website (C, D). Among them, we will recruit individuals with higher risk of amyloid positivity to participate in the J-TRC onsite study (E, F). J-TRC: Japanese Trial-Ready Cohort.



Methods

Ethics

The J-TRC web study was approved by the University of Tokyo Graduate School of Medicine Institutional Ethics Committee (ID: 2019132NI-(3)), and online informed consent was obtained from each participant upon registration for the analysis during collection of basic demographic information of the registrant. Upon using the J-TRC website, users are required to agree to the terms of use and the privacy policy of the website.

J-TRC Web Study Promotions

The details about the J-TRC web study were described in our recent report [26]. We were helped by Hakuhodo Inc [27] for the advertising and public relations of the J-TRC web study. The J-TRC web study has recruited more than 4000 online registrants from all over Japan within the first 9 months of its launch in October 2019. The promotion was conducted using multiple advertisement methods (Figure 1A), including a press release concurrent with the launch, advertisements in national newspapers, online advertisements, and direct email invitations to participants of another cognitive registry. These promotions were essentially performed one by one, but their timing interval, frequency, and duration period differed greatly, and they may have overlapped at some points. Financial incentives (eg, reward payments) [10] were not used. In this study, we reviewed the time-series data of promotions and the successive change in the target metrics during the consecutive period from October 31, 2019 to July 16, 2020 (for a total of 260 days).

After we conducted an initial press release simultaneously with the web study launch on October 31, 2019, the launch was reported by the media via newspapers or their websites or introduced by health-related web articles. This motivated target individuals to access the J-TRC web study website (we refer to this type of promotion as “press release reports” here). The J-TRC web study website is written in Japanese, and access to it is limited to domestic connection only. Since these press report activities were not advertisements and solely relied on the voluntary work of news media, we could not directly control the degree and range of the reports’ exposure in the media. Costs required for the initial press release, including the venue and equipment fee, was Japanese yen (JP ¥)7,620,000 in total. A currency exchange rate of US $1 = JP ¥108 was applicable on October 31, 2019. We confirmed the published date of newspaper reports or website articles introducing the J-TRC study using Nikkei Telecom [28], a large Japanese database covering newspapers, TV news, and magazines published in Japan. The impression data for each article were not available. Reports on radio news or other social media were not reviewed, as we were not able to access the reports on these media. The promotion press release reports were further separately treated, depending on the date of reports: whether it was reported within the first week (during which promotion activity was performed) or later, following the launch of the J-TRC web study.

Advertisements on the web were conducted in 2 ways: (1) listing advertising and (2) banner advertising. The listing advertisements (referred to as “web listing advertisements” here) used Google Ads [29] and Yahoo! Japan Ads [30] to display advertising texts recruiting J-TRC web study participants by specifically targeting internet users who were 60 years or older and searched for keywords associated with dementia, such as “物忘れ” (a Japanese word meaning “memory loss” or “memory impairment”) , “記憶障害” (ie, “memory loss” or “memory impairment”), “認知症” (ie, “dementia”) , “アルツハイマー病” (ie, “Alzheimer disease”) , “レビー小体型認知症” (ie, “dementia with Lewy Bodies”) , or “軽度認知障害” (ie, “mild cognitive impairment”) on the Google or Yahoo! search engines. Clicking the advertising sentence displayed on the search result directly connected the user to the J-TRC website (Figure 1C), which is counted as a onetime website visit. The listing advertising was conducted during the period

XSL•FO
RenderX

from December 12, 2019 to January 25, 2020 (for a total of 42 days), with a total budget of JP ¥4,200,000. There were 2,919,186 impressions, among which there were 24,024 clicks (click rate 0.82%) in total with Yahoo! Japan Ads (JP ¥114 per visitor). There were 428,104 impressions, among which there were 11,183 clicks (click rate 2.61%) in total with Google Ads (JP ¥123 per visitor).

Web banner advertising promotion (referred to as "web banner advertisement" here) was conducted during the period of Dec 12, 2019 to Mar 20, 2020 (for a total of 100 days), with a total budget of JP ¥1,468,000. Web banner advertisement, as exemplified in Figure 1B, was displayed on the dementia-related web article pages of Medical Note [31], a medical web media. Clicking the banner advertising also directly transferred the user to the J-TRC website (Figure 1C), as with the web listing advertisement. This advertisement received 286,231 impressions, of which 1,138 produced clicks (click rate 0.4%) in total (JP ¥1289 per visitor), as summarized in Table 1.

Table 1. Summary of costs required for each promotion method.

Promotion method	Total cost, JP ¥[a,b]	Total visitors to website	Actual cost per visitor, JP ¥[b]
Press release reports (1st and 2nd week+)	7,620,000	—[c]	—
Newspaper advertisement	6,300,000	—	—
Web listing advertisement	4,102,124	35,207	116.6
Web banner advertisement	1,468,000	1,138	1289
Email invitation to IROOP[d]	0[e]	—	—

[a]JP ¥: Japanese yen.

[b]A currency exchange rate of US $1 = JP ¥108 was applicable on October 31, 2019.

[c]—: not available.

[d]IROOP: Integrated Registry of Orange Plan.

[e]Sending the emails did not require any cost by itself.

Advertisements in the national newspaper (referred to as "newspaper advertisement" here) were conducted twice in total. The first ran on February 13th, 17th, and 19th of 2020 in The Yomiuri Shimbun [32] morning and evening editions, with a total cost of JP ¥2,300,000, with advertisements announcing the holding of a public seminar on the latest dementia research and the introduction of the J-TRC study. The second promotion ran on March 21, 2020 in The Yomiuri Shimbun morning edition, with a cost of JP ¥4,000,000 to advertise the J-TRC study directly. Advertisements in local and regional newspapers were not used.

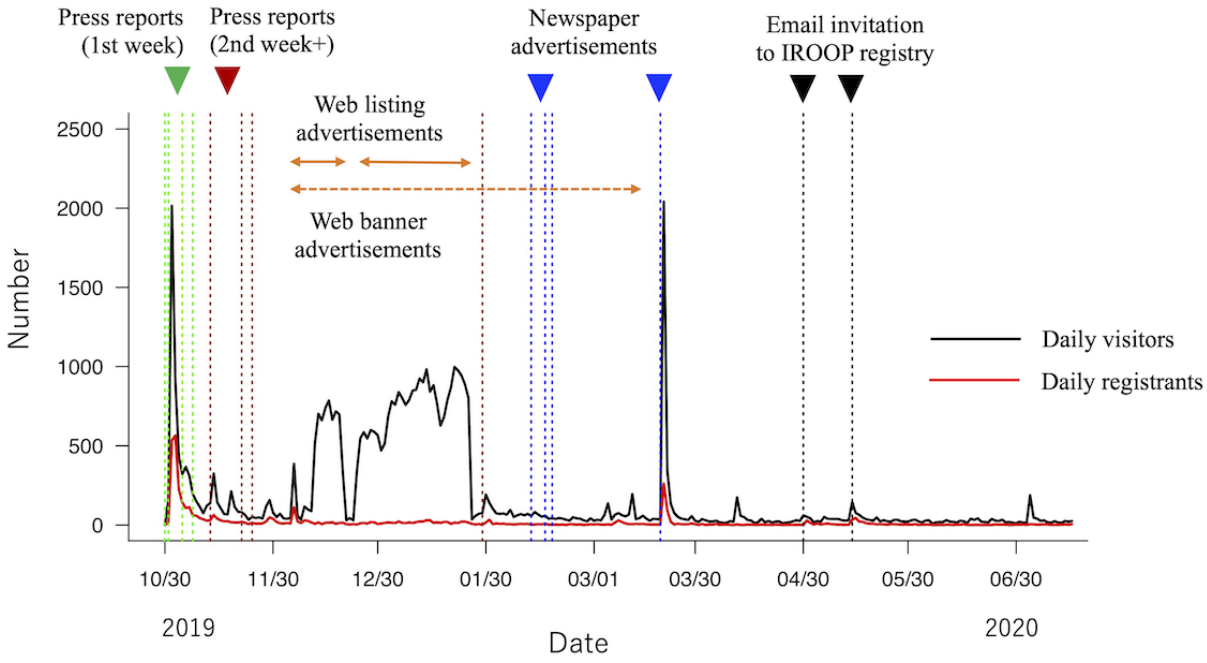
The direct email invitations were sent twice in total to participants in the Integrated Registry of Orange Plan (IROOP; referred to as "email invitation to IROOP" here) [33]. The IROOP is a multiple-layered cognitive registry aimed at developing a better understanding of dementia, improving dementia care management, and facilitating dementia clinical trials [33]. The IROOP participants receive regular cognitive tests and some useful information about dementia clinical studies. Sending the emails did not require any cost by itself.

Effectiveness Assessment of Promotions

As shown in Figure 2, the timing interval, frequency, and duration of each promotion method differed greatly: these promotions were conducted in a combined manner. Figure 2 shows a time-series increase in the number of daily visitors to the J-TRC web study website (black line) and the number of daily registrants to the J-TRC web study (red line). Exposure to each promotion method is represented by the vertical or horizontal lines: green lines correspond to the day when press reports were published within the first week after the study launch, brown lines correspond to the press reports published after the first week since the launch, blue lines correspond to the day of newspaper advertisements, and the black lines correspond to the day when email invitations to the IROOP registry were sent. The orange horizontal lines correspond to the period during which web listing advertisements were carried out, and the orange horizontal dotted line corresponds to the period during which web banner advertisements had been displayed.

XSL•FO
RenderX

Figure 2. Serial daily record of the number of web study website visitors and study registrants. Exposure dates for each promotion method are represented by the vertical or horizontal lines, as labeled. IROOP: Integrated Registry of Orange Plan.



We cannot always clearly separate which promotion method, by itself, led to increases in the target metrics (eg, daily number of visitors or registrants in Figure 2), so we used the vector-autoregression (VAR) model as a multivariate time-series analysis to avoid cross-contamination on the effects of different promotion methods. We used R package vars [34] to perform the VAR model calculation, and the adequate lag order was determined to be 1 based on the Akaike's Information Criteria. The following equations describe the VAR model used in this study with a lag order of 1:

$$Y_{1,t} = C_1 + \phi_{11}Y_{1,t-1} + \phi_{12}Y_{2,t-1} + \phi_{13}Y_{3,t-1} + \phi_{14}Y_{4,t-1} + \phi_{15}Y_{5,t-1} + \phi_{16}Y_{6,t-1} + \phi_{17}Y_{7,t-1} + \phi_{18}Y_{8,t-1} + \varepsilon_{1,t} \quad \textbf{(1)}$$

$$Y_{k,t} = C_k + \phi_{k1}Y_{1,t-1} + \phi_{k2}Y_{2,t-1} + \phi_{k3}Y_{3,t-1} + \phi_{k4}Y_{4,t-1} + \phi_{k5}Y_{5,t-1} + \phi_{k6}Y_{6,t-1} + \phi_{k7}Y_{7,t-1} + \phi_{k8}Y_{8,t-1} + \varepsilon_{k,t} \ (k = 2,3, \ldots 8) \quad \textbf{(2)}$$

Value Y_1 is one of the target metrics we aimed to evaluate, representing the number of daily new visitors to the J-TRC website (ie, "daily visitors" in this study), the number of daily new registrants to the J-TRC web study (ie, "daily registrants"), the daily registration rate among daily visitors (registration rate = daily visitors/registrants; ie, "daily conversion rate"), daily rate of eligible registrants to the J-TRC web study among all daily registrants (eligible registrant rate = eligible registrants/total registrants; ie, "daily eligibility rate"), or the median age of daily eligible registrants (ie, "registrants' age"). Daily data of these metrics and other eligible participants' demographics were extracted from the data server of the J-TRC web study, for which management is entrusted to Nittetsu Hitachi Systems Engineering, Inc [35]. Eligibility for participation in the J-TRC web study was defined as follows: participants who had completed the registration and demographic input, provided informed consent for study participation, had no prior history of a dementia or AD diagnosis, and were aged from 50 to 85 years old at the time of registration [26].

Values Y_2 through Y_7 are the daily binary exposure statuses (with or without, regardless of the degree of impressions) to each of the promotion modalities as follows: Y_2 stands for the initial press release reports within the first week since launch, Y_3 stands for the press release reports after the first week since launch, Y_4 stands for the national newspaper advertisements, Y_5 stands for the web listing advertisements, Y_6 stands for the web banner advertisements, and Y_7 stands for the email invitations to IROOP. In addition, we used Google Trends' [36] results to incorporate general nationwide trends [37] of interest in dementia in Japan, irrespective of the promotional activities of the J-TRC study. We allocated Y_8 to the daily trend in the relative search volume in Google Trends, where the maximum daily count of searches for "dementia" (in its corresponding Japanese keyword) within the included period is equal to 100%. Trends in social media such as Twitter were not incorporated, because the J-TRC web study and its related keywords had hardly ever been tweeted about (as confirmed in September 2020). As we had confirmed that a few series of variables, including Y_1-Y_8, were not stationary via the augmented Dickey-Fuller test with R package tseries [38], we did not calculate Granger causality. The obtained coefficients φ_{12} through φ_{18} in Equation 1 correspond to the effect sizes of change in the target metric Y_1 at 1 day following the application of each promotion modality (Y_2 through Y_8) on the previous day.

The coefficients of promotion methods were further used to derive cost-effectiveness (costs in JP ¥ required to achieve 1

XSL•FO
RenderX

J-TRC website visitor or 1 J-TRC web study registrant). Although the total number of impressions and clicks following the use of advertisements were available in the case of web advertising, such data were not available for press releases, newspaper advertisements, or email invitations. Therefore, we calculated the average cost-effectiveness of each promotion method using the following equation:

> Mean cost-effectiveness = (Total cost of 1 promotion method) / (Estimated coefficients of the method [for daily visitors or daily registrants] × Total days spent on that method) **(3)**

The cost-effectiveness of email invitations to the IROOP registry could not be evaluated adequately as it required no substantial costs.

Validation

We additionally conducted promotions by newspaper advertisements alone in February to March 2021, of which data we used to validate the effectiveness of newspaper advertisements as calculated above, thereby attempting to further reduce the risk of cross-contamination of different promotion methods. This was again examined by applying the VAR model, where included variables were limited to Y_4 and Y_8 in Equation 1 and Equation 2.

Statistical Analysis

All the above data handling and analyses were performed using R 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria). A *P* value <.05 was considered as significant if not otherwise mentioned. To summarize the continuous value, we used the mean and SD. To summarize the geographical distribution of eligible registrations across Japan, we used R packages choroplethr [39] and choroplethrAdmin1 [40].

Results

Overview of Serial Daily Records in the Target Metrics

During the review period from October 31, 2019 to June 17, 2020, there were 48,334 website visitors, 4429 registrations (eligible or not; 9.16% of all 48,334 website visitors), and 3081 eligible registrations (69.56% of all 4429 registrations). The basic demographics of the eligible participants were as follows: mean age was 62.3 (SD 8.8) years at registration, 54.79% (1688/3081) were female, and 40.28% (1241/3081) had a family history of dementia or AD. Prefecture-level geographical distribution of the number of eligible registrants across Japan is shown as a choropleth map with a scale of natural logarithms in Figure 1D. Regional differences in the number of eligible registrants are apparent, reflecting the basic population by prefecture.

Time series of the number of daily visitors to the website and the number of daily registrants to the J-TRC web study are plotted in Figure 2. By visual inspection, both the number of daily new visitors (black line) and the number of daily new registrants (red line) prominently increased during the first week, along with the initial press release reports (represented by green vertical lines). The number of daily visitors gradually decreased thereafter, occasionally showing a transient mild surge following the press release reports and web article publications. Not all the small surges correspond with the timings of the press release reports and web articles we annotated (represented by green or brown vertical lines), partly because we cannot capture all the published newspaper, web, or magazine articles by Nikkei Telecom. Web advertisements (listings or banners, represented by horizontal orange lines) greatly contributed to the number of daily visitors but seemed not to affect that of daily registrants. Newspaper advertisements (represented by blue vertical lines) significantly contributed to the transient increase in daily visitors (P<.001) and daily registrants (P=.001). Email invitations to IROOP registrants seemed to mildly increase the number of daily visitors. During the period when no promotions were conducted (eg, June 2020), the mean number of daily visitors was 23.4 (SD 8.5) visitors per day, and the mean number of daily registrants was 1.6 (SD 1.5) registrants per day.

Efficacy of Advertisements in Vector Autoregression Model

We quantitatively assessed the above visual inspection by multivariate time-series analysis using the VAR model. Table 2 presents the result coefficients and their adjusted *P* values in the VAR model with a lag order of 1. When we had employed 1 promotion method (shown in rows) over 1 day, the contribution of that method in changing the target metric (shown in columns) on the next day (with delay=1 day) is quantified as the coefficient value of the corresponding cell. For example, the press release reports at 1 day (within the first week) contributed to an increase of 716.6 in daily website visitors on the next day.

XSL•FO
RenderX

Table 2. Result coefficients of vector-autoregression model (2019-2020).

Promotion method	Y_1 (effect on):				
	Daily visitors	Daily registrants	Daily conversion rate	Daily eligibility rate	Registrants' age
Y_2 (press release reports, 1st week)					
Result coefficient	716.6	183.6	8.9	–8.4	0.8
P value	<.001	<.001	.06	.54	.84
Y_3 (press release reports, 2nd week+)					
Result coefficient	88.9	11.8	0	–7.5	–3.7
P value	.28	.52	.99	.53	.28
Y_4 (newspaper advertisements)					
Result coefficient	509.3	64	–2.2	–1.3	–2.4
P value	<.001	.001	.60	.92	.50
Y_5 (web listing advertisements)					
Result coefficient	486.6	6.7	–4.5	–14.8	–4.4
P value	.001	.38	.008	.003	.002
Y_6 (web banner advertisements)					
Result coefficient	–40.4	–6.2	–2.5	2.5	2.3
P value	.12	.29	.06	.50	.03
Y_7 (email invitations to IROOP[a])					
Result coefficient	–3.1	23.1	43.3	8.9	2.6
P value	.98	.38	<.001	.60	.59
Y_8 (trends in general interest in dementia, from Google Trends)					
Result coefficient	1.1	0.1	0	0.1	0
P value	.21	.52	.45	.55	.23

[a]IROOP: Integrated Registry of Orange Plan.

Overall, on average, press releases within the first week had the highest effect on increasing daily visitors (+716.6), followed by newspaper advertisements (+509.3), and the web listing advertisements (+486.6). The press releases within the first week had the highest effect on increasing the daily registrants (+183.6), followed by newspaper advertisements (+64). In addition, email invitations to IROOP registrants showed the highest effect on increasing the daily conversion rate (+43.3%), whereas web listing advertisements showed the opposite effect, lowering the daily conversion rate (–4.5%) and the daily eligibility rate (–14.8%). Furthermore, web banner advertisements contributed to making the median age of daily eligible registrants slightly older (+2.3 years old), while the web listing advertisement contributed to lowering the median age of daily eligible registrants (–4.4 years old). General trends of interest in the keyword “dementia” by the Japanese population, as represented by the relative search volume in Google Trends, had no effect on increasing nor decreasing the daily target metrics.

Cost-effectiveness Analysis

The average cost-effectiveness per website visitor was calculated using the VAR model coefficients (Table 2). Only promotion methods with a coefficient value higher than 0 were included in the calculation, because we did not assume a negative value in the number of visitors and registrants. The results are shown in Table 3. The results for web banner advertisements were negative and, therefore, not applicable.

XSL•FO
RenderX

Table 3. Cost-effectiveness analysis.

Promotion method	Total days of advertisement	Estimated visitors per 1 day of advertisement	Average cost per visitor, JP ¥[a,b]	Estimated registrants per 1 day of advertisement	Average cost per registrant, JP ¥[b]
Press release reports (1st week)	4	716.6	2658	183.6	10,376
Press release reports (2nd week+)	4	88.9[c]	21,429[c]	11.8[c]	161,440[c]
Newspaper advertisements	4	509.3	3093	64	24,609
Web listing advertisements	42	486.6	201	6.7[c]	14,592[c]
Web banner advertisements[d]	100	N/A[e]	N/A	N/A	N/A

[a]JP ¥: Japanese yen.

[b]A currency exchange rate of US $1 = JP ¥108 was applicable on October 31, 2019.

[c]The efficacy coefficient of these promotion methods were not significant (in Table 2), and, therefore, presented for reference only.

[d]The results for web banner advertisements were negative and therefore not applicable.

[e]N/A: Not applicable.

The press release reports within the first week and the newspaper advertisements had similar levels of cost-effectiveness, at approximately JP ¥3,000 per website visitor (JP ¥2,658 and JP ¥3,093, respectively; Table 3). The average cost of the web listing advertisements was calculated as JP ¥201 per visitor, which is within the same order as its direct cost-effectiveness (ie, JP ¥116.6; Table 1). In terms of the cost-effectiveness per registrant (Table 3), newspaper advertisements required the highest cost (JP ¥24,609 per registrant), followed by press release reports within the first week (JP ¥10,376 per registrant).

For reference, we also overviewed the cost-effectiveness of promotion methods, for which efficacy was not significant (*P* values shown in Table 2) in the current VAR model: the web listing advertisements had a similar level of cost-effectiveness for registrant recruitment (JP ¥14,592) as the press releases within the first week or as the newspaper advertisements. In addition, the press release reports in the second week or later had the worst level of cost-effectiveness (JP ¥21,429 per visitor and JP ¥161,440 per registrant).

Lastly, we validated the above results of newspaper advertisements in a period when there were no other promotions used, in 2021. Figure 3 shows the time series of the number of daily visitors to the website and the number of daily registrants to the J-TRC web study, just succeeding the data presented in Figure 2. During 2020, since April, we could not conduct any active promotions mainly due to the COVID-19 pandemic. Then, in February 2021, we decided to conduct promotions mainly by using newspaper advertisements, which we considered the most effective promotion method based on the results just described. In a period between the middle of February 2021 and the middle of March 2021, there were no other promotion methods conducted (Figure 3B), so that analyzing the serial changes in this period would allow us to see the pure effect of newspaper advertisements. The result coefficients of the VAR model are shown in Table 4. Although a bit smaller than the coefficients in Table 2, on average, newspaper advertisements showed a similar level of significant increases to the daily visitors (+295.2, *P*=.032) and daily registrants (+42.7, *P*=.025).

Calculated cost-effectiveness of newspaper advertisements, in this case, was JP ¥5226 per website visitor and JP ¥36,156 per registrant, which were only approximately 1.5× higher than the results shown in Table 3.

XSL•FO
RenderX

Figure 3. Serial daily record of the number of web study website visitors and study registrants, just succeeding the data presented in Figure 2. (A) Exposure dates for each promotion method are represented via vertical lines, as labeled. (B) More detailed exposure dates for the promotion method newspaper advertisements are represented via vertical lines. IROOP: Integrated Registry of Orange Plan.

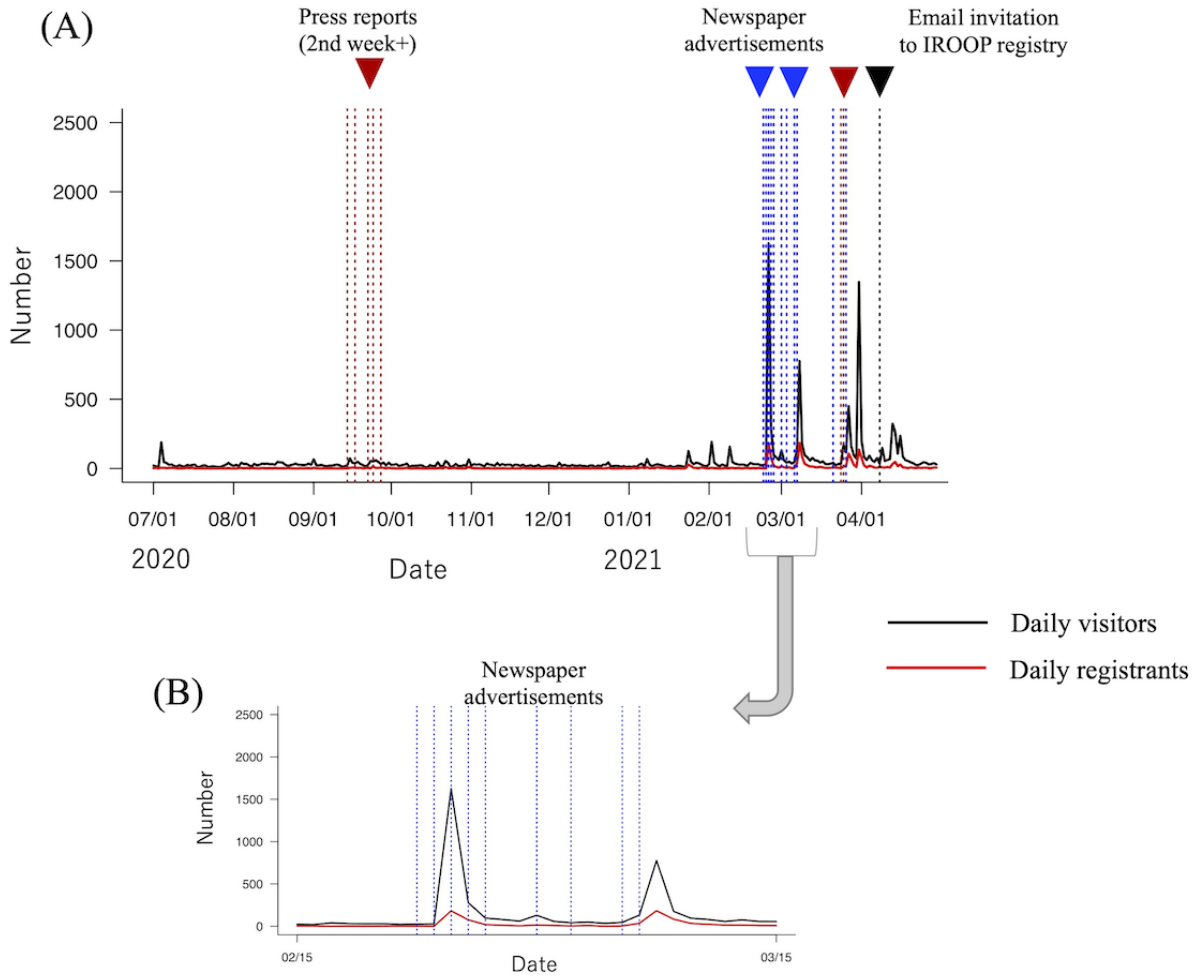


Table 4. Result coefficients of vector-autoregression model (2021).

Promotion method	Y_1 (effect on):				
	Daily visitors	Daily registrants	Daily conversion rate	Daily eligibility rate	Registrants' age
Y_4 (newspaper advertisements)					
Result coefficient	295.2	42.67	5.42	–9.65	0.4
P value	.03	.03	.13	.35	.90
Y_8 (trends in general interest in dementia, from Google Trends)					
Result coefficient	–1.64	–0.34	–0.1	–0.12	0.05
P value	.75	.63	.47	.77	.69

Discussion

Principal Findings

In this study, using multivariate time-series analyses, we quantitatively examined the recruitment efficacy of multiple types of promotion methods and their cost-effectiveness. Our results showed that each promotion modality had different features in their effects on website visitors and participant recruitment for the J-TRC web study. Under the advertising condition settings so far, the traditional methods of press releases and newspaper advertisements had a marked effect on increasing website visitors and study registrants with fair cost-effectiveness mostly equivalent to earlier studies while web-based advertisements also had a considerable effect on increasing website visitors with highly efficient costs. In addition, direct email invitations to other cognitive registries achieved the highest registration rate, and 2 web advertising methods altered the distribution of the registrants' age mildly and oppositely. These results can provide important suggestions for future

XSL•FO
RenderX

promotion for recruiting presymptomatic participants to AD clinical trials in Japan.

Although the promotion methods press release reports after the first week, web banner advertisements, or email invitations to IROOP registry seemed to have some effects on the number of daily visitors by visual inspection (Figure 2), these effects were not statistically significant. These results are those derived in the framework of the VAR model and under the advertisement settings so far, so careful interpretation is needed, as they may actually turn out to be effective in other statistical models or in other advertisement settings. For example, prolonging the period of web banner advertisements (eg, to 1 year or more, as in earlier studies [1]), changing the placement of the banner advertisement on the website, or using banner images with more visually appealing designs [11,41] may yield different results.

Similarly, these results do not always mean that web listing advertisements are not effective at recruiting registrations to the J-TRC web study. Because many earlier online studies have repeatedly reported the efficacy of web advertisements [3-6,8,9,11-13], much needs to be done to improve the efficacy of web listing advertisements in the J-TRC web study. The web listing advertisements contributed to increasing daily visitors greatly but did not greatly affect the daily registrants, suggesting a large proportion of visitors left the website without proceeding to registration. In addition, the web listing advertisements also decreased the rate of eligible participants among the registrants (Table 2), implying that website visitors via web listing advertisements may have been less eager to participate in the J-TRC web study, because many of our noneligible registrants had not fully completed their basic demographic information. Based on these points, it may be helpful to link the web advertisement to the landing web page, which motivates website visitors to participate in the web study.

The different promotion methods seemed to have recruited participants with slightly different backgrounds. For example, direct email invitations to other cognitive registries achieved the highest registration rate, presumably reflecting their high interest in the AD clinical studies [5]. We can, therefore, expect these registrants to have high eligibility as participants for the J-TRC study as well as successive AD clinical trials, which require years of continuous commitment from participants. In addition, 2 web advertising methods had incongruent effects on recruiting participants of slightly different ages, which we consider to be due to the difference in the age of the main users. Namely, the main users of the webpage where the web banner was displayed might have been slightly older than those who searched for dementia-related keywords by search engines. These features suggest that changing the allocation of promoting methods may help adjust the background of overall J-TRC study participants in a manner optimized for the screening cohort for preclinical AD.

We calculated the average cost-effectiveness by using coefficients in the VAR model, because data of daily visitors and registrants were not directly available, except for the data tracked by the web advertisements. The direct cost-effectiveness for web listing advertisements was JP ¥117 per visitor; the calculated average cost-effectiveness for the same web listing advertisements were similar, at JP ¥201 per visitor. This result supports the reliability of the average cost-effectiveness calculated in our study, suggesting we could compare the rough cost-effectiveness between the different promotion methods.

Traditional methods of press releases and newspaper advertisements were suspected to have the largest effect on increasing website visitors and study registrants, despite the considerable costs required. This may be due to the nature of newspapers: it was reported that the newspaper was considered as a highly important source of information and as the most reliable media by Japanese people in their 60s, when compared with television or internet [16]. The cost-effectiveness of newspaper advertisements was JP ¥24,609 per registrant, which is mostly equivalent to the costs reported in earlier studies recruiting via newspaper advertisements in Western countries, such as approximately UK £250 (US $322 on October 31, 2019) per eligible participant for a celecoxib clinical trial in Britain [14], Can $113 (US $86 on October 31, 2019) per enrolled participant for recruiting postmenopausal women in Canada [42], US $115 per recruited participant for a smoking cessation intervention [43], and Aus $239 (US $164 on October 31, 2019) per recruited eligible participant for elderly patients with diabetes in Australia [44]. Because the estimated efficacy of web advertisements to increase the J-TRC web study registrants turned out not to be significant (Table 2), we cannot compare the cost-effectiveness between newspaper advertisements and the web listing advertisements with confidence. However, the newspaper advertisements may also have a similar level of cost-effectiveness as web listing advertisements, which were reported to have a good cost-effectiveness in several earlier studies [3-6,8,9,11,12]. We also confirmed that the efficacy and cost-effectiveness of newspaper advertisements were largely stable whether they were measured in combination with other variables or measured alone with different timings (Table 2; Table 4). These points suggest that newspaper advertisements might be a reasonable method to promote registrations [13], even in the context of the J-TRC web study. The press release reports within the first week may also be effective to a certain extent, although we consider the generalizability of the effect of the press release reports to be rather limited because it cannot be conducted in replicate, unlike newspaper advertisements, nor can its degree of media coverage be controlled by researchers.

Limitations and Future Directions

Our study had several limitations. First, the VAR model presumes that the effect of each variable is fixed throughout the included period, but this is not always guaranteed: web advertisements or newspaper advertisements that were repeatedly displayed will inevitably lessen the recruitment effect. We also regarded all different media reports as a single category of press release reports and regarded the 2 different newspaper advertisement times (the first one for event announcement and the second one for the direct J-TRC promotion) as a single category of newspaper advertisements, which could lead to an inaccurate estimation in their average cost-effectiveness. In addition, the current VAR model–based approach to estimate the number of visitors and registrants may have underestimated the true number of visitors and registrants, because the estimated

XSL•FO
RenderX

efficacy of multiple promotions (as calculated by the sum of VAR significant coefficients × the total number of days of each promotion) was 25,340 visitors (among 48,334 true visitors) and 990 eligible registrants (among 3081 true eligible registrants). Although there is a similarity in the order of digits, this degree of underestimation in the efficacy of valid promotion methods suggests the true cost-effectiveness of the press releases and newspaper advertisements may be 2-3× lower than the estimated one reported here. This should be taken into consideration when planning future promotion strategies.

For complementing the shortness of this analysis, it may be helpful to follow up by determining the specific J-TRC web study promotions viewed by each participant. For example, a future study should make an online survey for website visitors or registrants asking how they learned of or were referred to the J-TRC study. However, it may not be easy to carry out within the current system, which was adopted from the Alzheimer Prevention Trials web study used in a similar preclinical AD study in the United States [25]. To use a data tracking system when sending emails to participants of other cognitive registries may be easier to implement.

In the future, we must address the effectiveness of social media such as Facebook, Twitter, Instagram, LINE, and TikTok to recruit elderly participants in the context of a Japanese clinical trial environment. Outside Japan, Facebook has been shown to be effective to recruit elderly participants to clinical trials [17,45]. In Japan, although the use rate of Facebook by elderly people has been relatively lower compared with younger people, it has been increasing in recent years. Facebook, Twitter, and Instagram were reported to have similar use rates, of approximately 10% in 2019 [16]; therefore, we can expect promotion via Facebook, Twitter, or Instagram might also be increasing in feasibility in Japan. Moreover, the use rate of elderly users of LINE, an instant communication app, dramatically increased between 2012 and 2019, and LINE is now the most popular social media platform in Japan (with approximately 70% use rate among people in their 60s in 2019) [16], meaning LINE may be one of the most useful media for promoting recruitment for the J-TRC web study, such as via web banner advertisements.

Conclusions

We quantitatively evaluated the efficacy and cost-effectiveness of multiple methods to promote the J-TRC web study. Our results showed that each promotion modality had different degrees of efficacy for recruiting visitors and participants to the web study website: under the advertisement setting conditions, traditional methods of initial press releases and newspaper advertisements were considered to have the largest efficacy in recruiting study registrants, with fair cost-effectiveness that was equivalent to earlier online studies of other disease domains, while web advertisements only contributed to increasing website visitors. These results can provide important suggestions for future promotion for recruiting presymptomatic participants to clinical trials of AD in Japan.

Acknowledgments

The authors are grateful to a team at Hakuhodo Inc and Nittetsu Hitachi Systems Engineering, Inc for their activities in promoting the J-TRC study. This study was supported by the Japan Agency for Medical Research and Development grants JP21dk0207048 and JP20dk0207028.

Conflicts of Interest

None declared.

References

1. Lane TS, Armin J, Gordon JS. Online Recruitment Methods for Web-Based and Mobile Health Studies: A Review of the Literature. J Med Internet Res 2015;17(7):e183 [FREE Full text] [doi: 10.2196/jmir.4359] [Medline: 26202991]
2. Perez T, Perez RL, Roman J. Conducting Clinical Research in the Era of Covid-19. Am J Med Sci 2020 Sep;360(3):213-215 [FREE Full text] [doi: 10.1016/j.amjms.2020.06.011] [Medline: 32690272]
3. Ramo DE, Prochaska JJ. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use. J Med Internet Res 2012;14(1):e28 [FREE Full text] [doi: 10.2196/jmir.1878] [Medline: 22360969]
4. Batterham PJ. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
5. Gaupp-Berghausen M, Raser E, Anaya-Boig E, Avila-Palencia I, de Nazelle A, Dons E, et al. Evaluation of Different Recruitment Methods: Longitudinal, Web-Based, Pan-European Physical Activity Through Sustainable Transport Approaches (PASTA) Project. J Med Internet Res 2019 Mar 09;21(5):e11492 [FREE Full text] [doi: 10.2196/11492] [Medline: 31066715]
6. Carter-Harris L, Bartlett ER, Warrick A, Rawl S. Beyond Traditional Newspaper Advertisement: Leveraging Facebook-Targeted Advertisement to Recruit Long-Term Smokers for Research. J Med Internet Res 2016;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
7. Thompson D, Canada A, Bhatt R, Davis J, Plesko L, Baranowski T, et al. eHealth recruitment challenges. Eval Program Plann 2006 Nov;29(4):433-440. [doi: 10.1016/j.evalprogplan.2006.08.004] [Medline: 17950873]
8. Graham AL, Milner P, Saul JE, Pfaff L. Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions. J Med Internet Res 2008;10(5):e50 [FREE Full text] [doi: 10.2196/jmir.1001] [Medline: 19073542]

XSL•FO
RenderX

9. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]
10. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrøm M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. BMJ Open 2013;3(2):1-24 [FREE Full text] [doi: 10.1136/bmjopen-2012-002360] [Medline: 23396504]
11. Ramo DE, Rodriguez TMS, Chavez K, Sommer MJ, Prochaska JJ. Facebook Recruitment of Young Adult Smokers for a Cessation Trial: Methods, Metrics, and Lessons Learned. Internet Interv 2014 Apr;1(2):58-64 [FREE Full text] [doi: 10.1016/j.invent.2014.05.001] [Medline: 25045624]
12. Loxton D, Harris ML, Forder P, Powers J, Townsend N, Byles J, et al. Factors Influencing Web-Based Survey Response for a Longitudinal Cohort of Young Women Born Between 1989 and 1995. J Med Internet Res 2019 Mar 25;21(3):e11286 [FREE Full text] [doi: 10.2196/11286] [Medline: 30907739]
13. Bracken K, Hague W, Keech A, Conway A, Handelsman DJ, Grossmann M, et al. Recruitment of men to a multi-centre diabetes prevention trial: an evaluation of traditional and online promotional strategies. Trials 2019 Jun 19;20(1):366 [FREE Full text] [doi: 10.1186/s13063-019-3485-2] [Medline: 31217024]
14. Hapca A, Jennings CG, Wei L, Wilson A, MacDonald TM, Mackenzie IS. Effectiveness of newspaper advertising for patient recruitment into a clinical trial. Br J Clin Pharmacol 2014 Jun;77(6):1064-1072 [FREE Full text] [doi: 10.1111/bcp.12262] [Medline: 24283948]
15. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
16. FY2020 Survey Report on Usage Time of Information and Communications Media and Information Behavior [article in Japanese]. URL: https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2020/9/30_04.html [accessed 2020-11-03]
17. Nash EL, Gilroy D, Srikusalanukul W, Abhayaratna WP, Stanton T, Mitchell G, et al. Facebook advertising for participant recruitment into a blood pressure clinical trial. J Hypertens 2017 Dec;35(12):2527-2531. [doi: 10.1097/HJH.0000000000001477] [Medline: 28704263]
18. Juraschek SP, Plante TB, Charleston J, Miller ER, Yeh H, Appel LJ, et al. Use of online recruitment strategies in a randomized trial of cancer survivors. Clin Trials 2018 Dec;15(2):130-138 [FREE Full text] [doi: 10.1177/1740774517745829] [Medline: 29361843]
19. Langbaum JB, Karlawish J, Roberts JS, Wood EM, Bradbury A, High N, et al. GeneMatch: A novel recruitment registry using at-home APOE genotyping to enhance referrals to Alzheimer's prevention studies. Alzheimers Dement 2019 Apr;15(4):515-524 [FREE Full text] [doi: 10.1016/j.jalz.2018.12.007] [Medline: 30772251]
20. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):280-292 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.003] [Medline: 21514248]
21. Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013 Feb;12(2):207-216 [FREE Full text] [doi: 10.1016/S1474-4422(12)70291-0] [Medline: 23332364]
22. Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, Australian Imaging, Biomarkers,Lifestyle Flagship Study of Ageing, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Cooperative Study. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. JAMA Neurol 2014 Aug;71(8):961-970 [FREE Full text] [doi: 10.1001/jamaneurol.2014.803] [Medline: 24886908]
23. Cummings J. The National Institute on Aging-Alzheimer's Association Framework on Alzheimer's disease: Application to clinical trials. Alzheimers Dement 2019 Jan;15(1):172-178 [FREE Full text] [doi: 10.1016/j.jalz.2018.05.006] [Medline: 29936146]
24. Japanese Trial-Ready Cohort (J-TRC) web study. 2019. URL: https://www.j-trc.org/ [accessed 2020-11-03]
25. Aisen PS, Sperling RA, Cummings J, Donohue MC, Langford O, Jimenez-Maggiora GA, et al. The Trial-Ready Cohort for Preclinical/Prodromal Alzheimer's Disease (TRC-PAD) Project: An Overview. J Prev Alzheimers Dis 2020;7(4):208-212 [FREE Full text] [doi: 10.14283/jpad.2020.45] [Medline: 32920621]
26. Sato K, Ihara R, Suzuki K, Niimi Y, Toda T, Jimenez-Maggiora G, et al. Predicting amyloid risk by machine learning algorithms based on the A4 screen data: Application to the Japanese Trial-Ready Cohort study. Alzheimers Dement (N Y) 2021;7(1):e12135 [FREE Full text] [doi: 10.1002/trc2.12135] [Medline: 33778148]
27. Hakuhodo Website. URL: https://www.hakuhodo-global.com [accessed 2020-11-03]
28. Nikkei Telecom. URL: http://telecom.nikkei.co.jp [accessed 2020-11-03]
29. Google Ads. URL: https://ads.google.com/intl/ja_jp/home/ [accessed 2020-11-03]
30. Yahoo! Japan Ads. URL: https://promotionalads.yahoo.co.jp [accessed 2020-11-03]
31. Medical Note. URL: https://medicalnote.jp [accessed 2020-11-03]
32. Yomiuri Shimbun. URL: https://www.yomiuri.co.jp [accessed 2020-11-03]

XSL•FO
RenderX

33. Saji N, Sakurai T, Suzuki K, Mizusawa H, Toba K, ORANGE investigators. ORANGE's challenge: developing wide-ranging dementia research in Japan. Lancet Neurol 2016 Jun;15(7):661-662. [doi: 10.1016/S1474-4422(16)30009-6] [Medline: 27302232]
34. Pfaff B. VAR, SVAR and SVEC Models: Implementation Within R Package vars. Journal of Statistic Software. URL: http://www.jstatsoft.org/v27/i04/ [accessed 2020-11-03]
35. Nittetsu Hitachi Systems. URL: https://www.nhs.co.jp/en/ [accessed 2020-11-03]
36. Google Trends. URL: https://trends.google.co.jp/trends [accessed 2020-11-03]
37. Gualano MR, Lo Moro G, Voglino G, Bert F, Siliquini R. Monitoring the impact of COVID-19 pandemic on mental health: a public health challenge? Reflection on Italian data. Soc Psychiatry Psychiatr Epidemiol 2020 Oct 09:165-167 [FREE Full text] [doi: 10.1007/s00127-020-01971-0] [Medline: 33034669]
38. Trapletti A, Hornik K. tseries: Time Series Analysis and Computational Finance. R package version 0. URL: https://cran.r-project.org/web/packages/tseries/index.html [accessed 2020-11-03]
39. Lamstein A. choroplethr: Simplify the Creation of Choropleth Maps in R. R package version 3. URL: https://CRAN.R-project.org/package=choroplethr [accessed 2020-11-03]
40. Lamstein A. choroplethrAdmin1: Contains an Administrative-Level-1 Map of the World. R package version 1. URL: https://CRAN.R-project.org/package=choroplethrAdmin1 [accessed 2020-11-03]
41. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
42. Butt DA, Lock M, Harvey BJ. Effective and cost-effective clinical trial recruitment strategies for postmenopausal women in a community-based, primary care setting. Contemp Clin Trials 2010 Sep;31(5):447-456. [doi: 10.1016/j.cct.2010.06.003] [Medline: 20601159]
43. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. Nicotine Tob Res 2006 Dec;8 Suppl 1:S35-S41. [doi: 10.1080/14622200601039014] [Medline: 17491169]
44. Miller EG, Nowson CA, Dunstan DW, Kerr DA, Solah V, Menzies D, et al. Recruitment of older adults with type 2 diabetes into a community-based exercise and nutrition randomised controlled trial. Trials 2016 Sep 26;17(1):467 [FREE Full text] [doi: 10.1186/s13063-016-1589-5] [Medline: 27669823]
45. Cowie JM, Gurney ME. The Use of Facebook Advertising to Recruit Healthy Elderly People for a Clinical Trial: Baseline Metrics. JMIR Res Protoc 2018 Jan 24;7(1):e20 [FREE Full text] [doi: 10.2196/resprot.7918] [Medline: 29367186]

Abbreviations

AD: Alzheimer disease
IROOP: Integrated Registry of Orange Plan
JP ¥: Japanese yen
J-TRC: Japanese Trial-Ready Cohort
VAR: vector-autoregression

Edited by R Kukafka; submitted 05.12.20; peer-reviewed by K Reuter, B Woo; comments to author 24.01.21; revised version received 03.02.21; accepted 24.05.21; published 22.07.21.

Please cite as:
Sato K, Niimi Y, Ihara R, Suzuki K, Toda T, Iwata A, Iwatsubo T
Efficacy and Cost-effectiveness of Promotion Methods to Recruit Participants to an Online Screening Registry for Alzheimer Disease Prevention Trials: Observational Study
J Med Internet Res 2021;23(7):e26284
URL: https://www.jmir.org/2021/7/e26284
doi:10.2196/26284
PMID:34292159

©Kenichiro Sato, Yoshiki Niimi, Ryoko Ihara, Kazushi Suzuki, Tatsushi Toda, Atsushi Iwata, Takeshi Iwatsubo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Health Information Quality Assessment Tool for Korean Online Newspaper Articles: Development Study

Naae Lee[1], MPH; Seung-Won Oh[2,3*], MD, MBA, PhD; Belong Cho[3,4*], MD, PhD; Seung-Kwon Myung[5,6,7], MD, PhD; Seung-Sik Hwang[1], MD, PhD; Goo Hyeon Yoon[8], BA

[1]Department of Public Health Science, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

[2]Department of Family Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea

[3]Department of Family Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[4]Department of Family Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[5]Department of Cancer Biomedical Science, National Cancer Center Graduate School of Cancer Science and Policy, Goyang, Republic of Korea

[6]Division of Cancer Epidemiology and Management, Research Institute, National Cancer Center, Goyang, Republic of Korea

[7]Department of Family Medicine and Center for Cancer Prevention and Detection, Hospital, National Cancer Center, Goyang, Republic of Korea

[8]Liver Korea, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Seung-Won Oh, MD, MBA, PhD
Department of Family Medicine
Healthcare System Gangnam Center
Seoul National University Hospital
38-40FL Gangnam Finance Center 152, Teheran-ro, Gangnam-gu
Seoul, 06236
Republic of Korea
Phone: 82 2 2112 5500
Email: sw.oh@snu.ac.kr

Abstract

Background: Concern regarding the reliability and accuracy of the health-related information provided by online newspaper articles has increased. Numerous criteria and items have been proposed and published regarding the quality assessment of online information, but there is no standard quality assessment tool available for online newspapers.

Objective: This study aimed to develop the Health Information Quality Assessment Tool (HIQUAL) for online newspaper articles.

Methods: We reviewed previous health information quality assessment tools and related studies and accordingly developed and customized new criteria. The interrater agreement for the new assessment tool was assessed for 3 newspaper articles on different subjects (colorectal cancer, obesity genetic testing, and hypertension diagnostic criteria) using the Fleiss κ and Gwet agreement coefficient. To compare the quality scores generated by each pair of tools, convergent validity was measured using the Kendall τ ranked correlation.

Results: Overall, the HIQUAL for newspaper articles comprised 10 items across 5 domains: reliability, usefulness, understandability, sufficiency, and transparency. The interrater agreement for the article on colorectal cancer was in the moderate to substantial range (Fleiss κ=0.48, SE 0.11; Gwet agreement coefficient=0.74, SE 0.13), while for the article introducing obesity genetic testing it was in the substantial range, with values of 0.63 (SE 0.28) and 0.86 (SE 0.10) for the two measures, respectively. There was relatively low agreement for the article on hypertension diagnostic criteria at 0.20 (SE 0.10) and 0.75 (SE 0.13), respectively. Validity of the correlation assessed with the Kendall τ showed good correlation between tools (HIQUAL vs DISCERN=0.72, HIQUAL vs QUEST [Quality Evaluation Scoring Tool]=0.69).

Conclusions: We developed a new assessment tool to evaluate the quality of health information in online newspaper articles, to help consumers discern accurate sources of health information. The HIQUAL can help increase the accuracy and quality of online health information in Korea.

(J Med Internet Res 2021;23(7):e24436) doi:10.2196/24436

XSL•FO
RenderX

KEYWORDS

assessment tools; information seeking; newspaper articles; online health information; quality assessment

Introduction

With advancements in technology, public interest toward health has increased. This has led to the public actively seeking health-related information and enhancing their medical expertise beyond simply managing their diseases [1], which has had a positive impact on health-related behaviors and beliefs [2]. Unfortunately, validating the accuracy of information can be difficult because there is enormous asymmetry of health-related information among providers and consumers [3]. The asymmetry of this information further creates a gap between consumers, expressed through the consumers' health literacy or the production and distribution of inaccurate information [4,5]. To mediate this gap, the assessment of the quality of health-related information and the subsequent provision of the results to both providers and consumers must be undertaken using a standardized assessment tool. This will allow consumers to identify reliable information and reduce the risk of distributing channels of inaccurate health-related information [6,7].

There are various tools used to assess the quality of health-related information, including the DISCERN instrument, created by the University of Oxford [8]; the Health on the Net Foundation Code of Conduct (HONcode), developed by the Health on the Net Foundation in Switzerland [9]; and MedCERTAIN, supported by the Action Plan for Safer Use of the Internet of the European Union [10]. In the Republic of Korea, several tools to assess the quality of health-related information on the internet have also been developed [11-13]. Despite this, some of these instruments have not been designed to evaluate the quality of the information. Most of the tools do not evaluate online health information from newspaper articles, and their validity and reliability have not been verified [7,9,11,13]. Furthermore, these tools have a specific targeted format, thereby making it difficult to apply them to other media [12]. There are no gold-standard quality assessment tools for online health information [14]. DISCERN is a proven tool for validity and interrater reliability, but the validity and reliability of the Korean version has not been confirmed. Additionally, it is limited in the scope of application, as it is focused only on treatment information and is not applicable to online content about other aspects of health and illness, such as prevention and diagnosis, commonly covered in newspapers [15]. QUEST (Quality Evaluation Scoring Tool), which was recently developed for evaluating online health information, has also been proven to be valid, but it has not yet been translated into Korean [6], which makes it difficult to evaluate online health information in Korea.

With the increased use of the internet for information dissemination, the numbers of online newspaper articles and users have rapidly increased. According to a survey conducted in 2018, 63% of 3425 participants indicated that they preferred using the web to receive news, while only 17% of them chose print newspapers [16]. The results were particularly pronounced among younger generations. In 2016, the Nielsen Scarborough study noted that 49% of people accessed the internet to read newspaper articles in digital form instead of print [17].

According to the Korea Press Foundation, this trend can be seen in Korea as well, where 80.8% of the total population read newspaper articles via a mobile device [18]. As the number of online newspaper articles featuring health-related content has increased [19], there has been a rise in the number of problems stemming from inaccurate articles. This has created a need for addressing the quality assessment and management of health-related information [20]. A study that analyzed press reports on depression found that one-third of the articles did not mention the causes of depression at all, and only about half of the articles mentioned treatment methods [21]. Another study analyzed newspaper articles about sterility and found that most of them described infertile couples as abnormal or incomplete, consequently strengthening social prejudices [22]. Recently, it was reported that inaccurate newspaper articles can cause confusion among consumers when they are disseminated via social media [23]. Accordingly, the Association of Health Care Journalists suggested some fundamental principles to be followed when writing health-related articles, including professionalism, content, accuracy, independence, integrity, and responsibility [17]. Additionally, HealthNewsReview.org [24], a website that evaluates the quality of medical-related newspaper articles, has failed to describe in detail the process it adopts for developing criteria, and it has shown no evidence for individual criteria. Although the problems caused by health-related articles have increased, there are no suitable quality assessment tools to evaluate the quality of health-related newspaper articles in Korea. Therefore, this study aimed to develop the Health Information Quality Assessment Tool (HIQUAL) to assess health information in online newspaper articles.

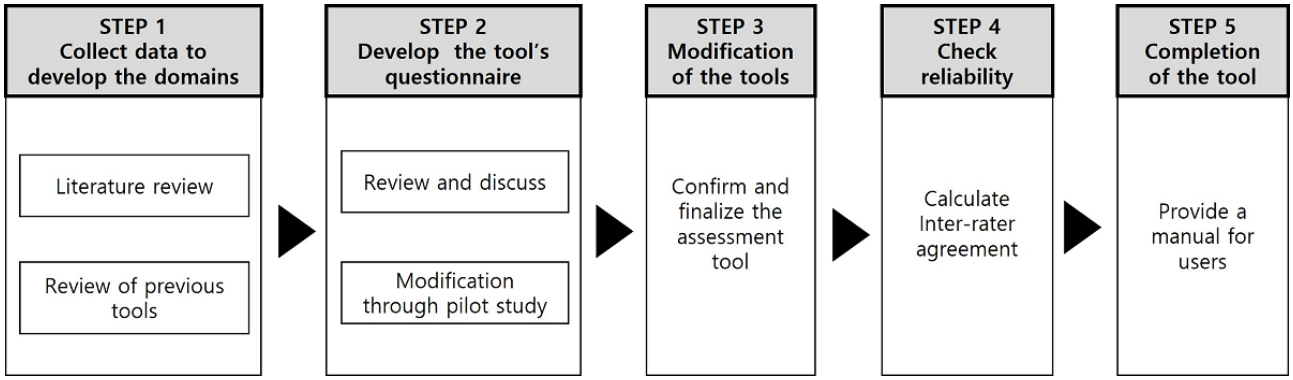
Methods

Overview

This study can be divided into four steps (Figure 1). First, we reviewed previous literature on the evaluation of health information quality assessment tools to develop the evaluation indicators. Second, we developed a draft of domains and questions through meetings and preliminary evaluations. Third, the assessment tool was modified and confirmed through evaluations and reviews at two different points in time. Fourth, we concluded the final agreement and validity with the assessment tool. The tool developed in this study—HIQUAL—was funded by the Korean Medical Association research project, which aims to develop standardized assessment tools and methods for systematic evaluation of health information from newspaper articles, television, and books.

XSL•FO
RenderX

Figure 1. The process of developing the tool.



Review of Previous Health Information Quality Assessment Tools

A review of existing literature was conducted to select the domains that correspond to the content of the questions guiding the development of the HIQUAL. A study by Wang and Strong was used to classify various dimensions included in existing assessment tools [25]. In this study, the data quality (DQ) dimension was divided into several domains—*intrinsic*, *contextual*, *representational*, and *accessibility*—and presented in a hierarchical framework to understand DQ from a consumer's perspective. Intrinsic DQ means that facts have quality in their own right; contextual DQ emphasizes the situations that must be considered to assess the context of the information; representational DQ emphasizes a format that is concise and consistent and data whose meaning is understandable and interpretable; and accessibility DQ emphasizes the significance of the parts of the framework. Our study also used this category to organize domains of various previously developed evaluation tools.

Developed by Oxford University, the DISCERN instrument evaluates information on disease and treatment under 3 domains, including 8 reliability items, 7 quality of information items, and 1 comprehensive evaluation, and has established feasibility and reliability [8,25,26]. The HONcode, developed by the Health on the Net Foundation, offers 8 ethical codes that health information websites must follow: authority, complementarity, confidentiality, attribution, justifiability, transparency, financial disclosure, and advertising [9]. The American Medical Association provides guidelines for health information websites using 4 categories: content, advertising and sponsorship in online posting, privacy and confidentiality of site visitors, and effectiveness and security of e-commerce [27]. MedCERTAIN is a third-party certification system developed as part of a project supported by the European Union's Action Plan for Safer Use of the Internet. The assessment items consist of identification, feedback, operation, and site identification of "information providers," as well as content, disclosure, policy, service, accessibility, and quality [10].

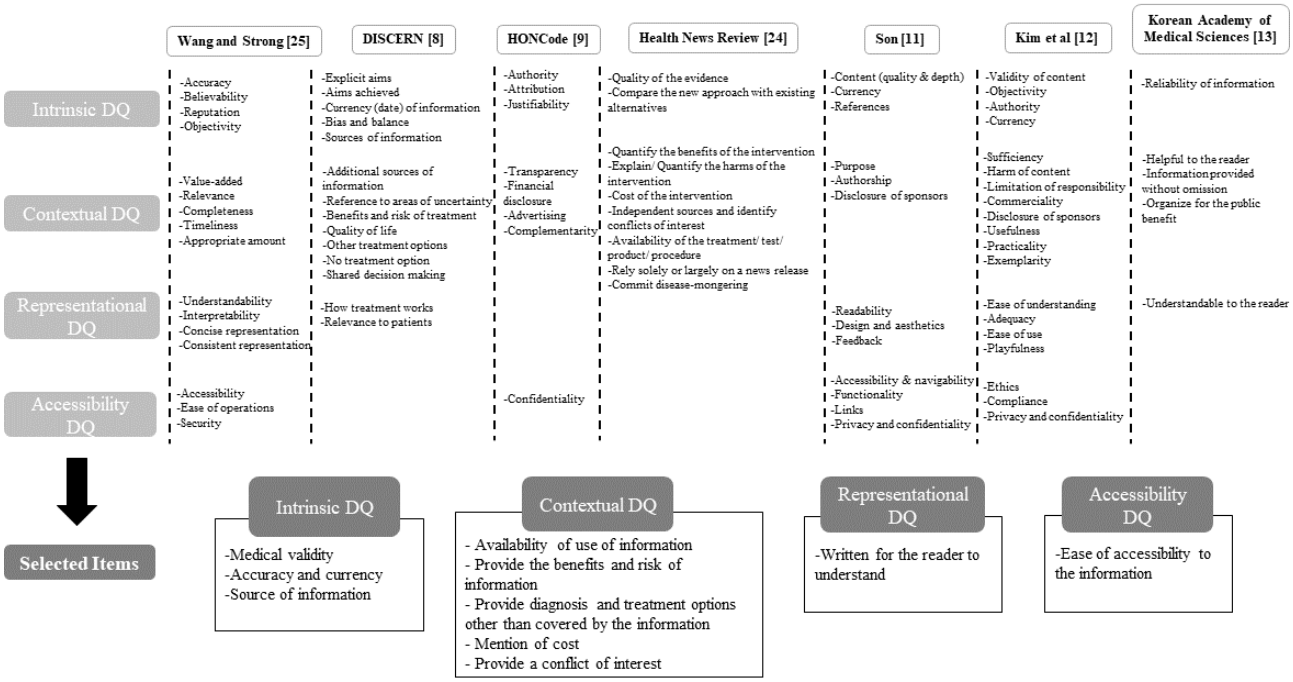
In Korea, Lee et al classified the common domains presented in previous studies as representation, contents, usage, and connection [28]. A study by Son referred to prior research and presented the evaluation criteria for health-related information websites in 7 domains: quality of content, authorship, purpose, design and aesthetic, functionality, contact address and feedback mechanism, and privacy [11]. The Internet Health Information Quality Checklist developed by Kim et al is characterized by different questions, depending on the user [12]. For professional use, it presents specific questions for items such as validity, sufficiency, and harmfulness of content, and the management is divided into purpose, authority, clarity of sponsorship, limitations of timing, commerciality, responsibility, privacy and security, ethics, and compliance. For general use, the question categories are divided into ease of understanding, adequacy, usefulness, sufficiency, appositeness, timeliness, admonition, ease of use, informativeness, and amusement [12]. The Korean Academy of Medical Sciences selected 5 evaluation criteria—reliability, usefulness, understandability, completeness, and publicity—and uses them for health information review certification projects [13]. The Health Information Monitoring Project of the Korean Medical Association uses a set of 6 criteria: scientific soundness, usefulness, sufficiency of information, whether facts are exaggerated, ease of use, and advertising. The Korea Institute for Health and Social Affairs conducted an evaluation of internet health information based on a set of 8 criteria: purpose (obviousness), appropriateness, accuracy, reliability, ease of use, authority, communication, and persistence [29].

Development of the Draft of the HIQUAL

The dimensions of the existing tools were classified according to the categories used by Wang and Strong (Figure 2) [25]. The indicators were assessed for the importance given to quality in online newspapers, providing the basis for the indicators and domains to be included in the new evaluation tool. Since the subject of the assessment tool was newspaper articles that are available in portal websites, the domain corresponding to accessibility was excluded, while draft questions were created for the other 3 domains. A rater with expertise in preventive medicine conducted a preliminary evaluation of 10 online health-related newspaper articles with draft questions for the newly developed HIQUAL. In this process, we compared the strengths and weaknesses of the existing evaluation tools with the newly selected questions.

XSL•FO
RenderX

Figure 2. Selection of the domain through comparison of existing evaluation tools. DQ: data quality; HONCode: Health on the Net Foundation Code of Conduct.



Completion of the Final Tool and Evaluation of Reliability and Validity

We analyzed the interrater agreement to ensure the consistency of the evaluation tool. A total of 3 interrater agreement analyses were performed in the process of revising the HIQUAL. The first and second evaluations were performed in a pilot study using a nonfinalized version of the tool with 3 articles, while the third evaluation was performed using the final version of the tool. News articles used in the assessment were randomly selected from among the most-viewed articles of the month in the health section of a portal website, which was the most popular online search engine in Korea [30]; the final evaluation was conducted in March 2018. In the first evaluation, 6 raters (3 family medicine physicians, 2 internists, and 1 obstetrician) participated [31-33]; in the second evaluation, 14 raters (8 family medicine physicians, 1 preventive medicine physician, 4 family medicine residents, and 1 representative of a patients' organization) participated. The second evaluation was conducted after a brief face-to-face training about the HIQUAL, and the interrater agreement was confirmed for the 2 evaluations. In the case of items with low agreement, discrepancies were identified, and the questions were revised by considering the opinions of the raters during the evaluation process. Finally, the third evaluation was conducted using the final version of the HIQUAL, and 5 raters (3 family medicine physicians, 1 preventive medicine physician, and 1 representative of a patients' organization) who had participated in the previous evaluations reviewed 3 new articles to reach an agreement.

The interrater agreement used two methods (Fleiss κ coefficient and Gwet agreement coefficient [AC]). Fleiss κ is a method used to measure the degree of agreement between two or more raters [34,35], where higher values indicate greater agreement. Gwet AC has the advantage of being able to accurately estimate population values without responding to ambient probabilities by taking changes between raters into account [36,37] and not being vulnerable to the kappa paradox [38]. The κ statistic tends to have a low value although there is strong interrater agreement; this can lead to kappa paradox and produce a biased result [39]. Gwet AC overcomes the κ limitation since it provides a stable interrater agreement and is less affected by prevalence and marginal probability; thus, it is used as a "paradox-resistant" alternative interrater coefficient [39]. Benchmarking scales of the Fleiss scale, AC statistics of 0.40 or less indicate poor agreement, 0.40-0.75 indicates a moderate to good agreement, and 0.75 or higher indicates excellent agreement [35,37,39]. For interrater agreement, the "kappaetc" package was used with the Stata version 16 (StataCorp) statistical software program.

The validity was verified by comparing the results of 3 tools: DISCERN, QUEST, and HIQUAL. The Kendall rank correlation coefficient (τ) proposed by Kendall measures the association and strength between paired observations [40]. The Kendall τ has better statistical properties of distribution [41] compared to other rank correlations and is easy to calculate [40]. We evaluated the 16 articles that were used in the preliminary evaluation and reliability analysis by one rater, using the DISCERN, QUEST, and HIQUAL tools. We divided the articles into 3 categories: treatment-related, diagnosis-related, and prevention-related; a total of 9 correlational tests were performed. The Kendall τ correlation coefficient returns a value of 0 to 1, where 0 indicates no relationship and 1 indicates perfect relationship. As a rule of thumb, the strengths of the correlation categories are as follows: 0.00 to 0.30 (0.00 to −0.30) indicates a negligible correlation, 0.30 to 0.50 (−0.30 to −0.50) indicates a low positive (negative) correlation, 0.50 to 0.70 (−0.50 to −0.70) indicates a moderate positive (negative) correlation, 0.70 to 090 (−0.50 to −0.70) indicates a high positive (negative) correlation, and 0.90 to 1.00 (−0.90 to −1.00) indicates a very high positive (negative) correlation [42]. For

XSL•FO
RenderX

the validity tests, we used Stata version 16 and SPSS version 26 (IBM Corp).

Results

Tool Overview

The final version of the HIQUAL is presented in a table format. The newly developed tool consists of the five domains of reliability, usefulness, understandability, sufficiency, and transparency, and has 10 questions (Figure 3 and Multimedia Appendix 1). The evaluation results are divided into 3 categories—Yes (1 point), No (0 points), and Not Applicable (NA)—and the final score is calculated by adding up the corresponding scores. NA results are excluded when calculating the total score. For example, if out of the 10 questions, 7 are marked as "yes," 2 as "no," and 1 as "NA," the total score is 7 out of 9 points (78%), instead of 7 out of 10 points (70%).

Figure 3. Health Information Quality Assessment Tool (HIQUAL) for health-related newspaper articles (English translated version).

Title:		Publisher:	Published: 20..
Domain	Item	Score	Comment
Reliability	1. It is medically valid and written based on scientific evidence.	0/1	
	2. The source of the information was presented.	0/1	
	3. It presented the timing of the information and is up to date.	0/1	
	4. There are no errors or exaggerations in the article.	0/1	
Usefulness	5. It is helpful to readers and explained the availability of information.	0/1	
Understandability	6. It was explained in a way the reader could understand.	0/1	
Sufficiency	7. The benefits and harms associated with target information (health risk factors, diagnosis, treatment, etc) are presented together.	0/1/NA	
	8. The target information (health risk factors, diagnosis, treatment, etc) was compared to other existing alternatives.	0/1/NA	
	9. The costs associated with target information (health risk factors, diagnosis, treatment, etc) were described.	0/1/NA	
Transparency	10. If there is no interest or if there is a conflict of interest, it was presented.	0/1	
Total score		Rater	Evaluation date
/			20..

Reliability Analysis

Using the HIQUAL, 5 raters evaluated 3 online newspaper articles. The interrater agreement was then analyzed (Table 1). Agreement was in the moderate to substantial range for the colorectal cancer–related article [43] (Fleiss κ=0.49, SE 0.11; Gwet AC=0.74, SE 0.13) and in the substantial range for the article introducing obesity genetic testing [44] (Fleiss κ=0.63, SE 0.28; Gwet AC=0.86, SE 0.10). In contrast, the article introducing the study of the changed hypertension diagnostic criteria showed a low level of agreement for Fleiss κ at 0.20 (SE 0.10) but a substantial agreement for Gwet AC at 0.75 (SE 0.13) [45]. For this article, the results of 4 raters' evaluations, excluding 1 representative of a patients' organization, showed a moderate agreement with a value of 0.40 (SE 0.19) for Fleiss κ, while showing an excellent agreement for Gwet AC at 0.85 (SE 0.11). With a reanalysis of the evaluation results, we confirmed that the level of agreement increased when we examined the results of the 4 medical specialists and excluded those of the nonmedical rater.

In terms of the statistical values of agreement, the coefficients for Fleiss κ were lower than those for Gwet AC. However, the overall percentage of agreements, including that of the third article with the lowest interrater agreement, was higher than 0.70. Fleiss κ is a model developed in Cohen κ, which may show the kappa paradox [35]. Gwet AC provides a more stable agreement than κ [39]; hence, in this case, it may be more appropriate to select the Gwet AC statistic [38,39]. For the value of Gwet AC, all 3 articles showed a high, close to excellent agreement.

Table 1. The interrater agreement based on 3 articles with 5 raters.

Statistic	Coefficient (SE)		
	Article 1	Article 2	Article 3
Scott/Fleiss κ	0.49 (0.11)	0.63 (0.28)	0.20 (0.10)
Gwet agreement coefficient	0.74 (0.13)	0.86 (0.10)	0.75 (0.13)
Percentage of agreement	0.79 (0.09)	0.90 (0.07)	0.78 (0.10)

Validity Analysis

The results of the validity tests are shown in Table 2. Sixteen online newspaper articles evaluated using Kendall τ showed a moderate to high correlation between the tools. For the 16 articles as a whole, the lowest correlation was obtained when comparing HIQUAL to QUEST, with the lowest at 0.69 and the highest at 0.72; there was a strong correlation when comparing HIQUAL and DISCERN. With treatment-related articles, the comparison between HIQUAL and QUEST was the lowest at 0.59, and the comparison between QUEST and DISCERN showed the highest correlation at 0.75. The lowest correlation was between QUEST and DISCERN (0.48) for the articles with content on topics other than treatment, and the highest correlation was between HIQUAL and DISCERN (0.67).

XSL•FO
RenderX

Table 2. Validity test with Kendall τ, SE, 95% CI, and *P* value of each test for health-related articles.

Article category	Kendall τ (95% CI)
Total articles (n=16)	
HIQUAL[a] vs DISCERN	0.72 (0.49-0.86)
HIQUAL vs QUEST[b]	0.69 (0.44-0.84)
QUEST vs DISCERN	0.70 (0.45-0.84)
Treatment articles (n=7)	
HIQUAL vs DISCERN	0.62 (0-0.90)
HIQUAL vs QUEST	0.59 (0-0.89)
QUEST vs DISCERN	0.75 (0.22-0.94)
Diagnosis and prevention articles (n=9)	
HIQUAL vs DISCERN	0.67 (0.22-0.88)
HIQUAL vs QUEST	0.65 (0.19-0.87)
QUEST vs DISCERN	0.48 (0-0.80)

[a]HIQUAL: Health Information Quality Assessment Tool.

[b]QUEST: Quality Evaluation Scoring Tool.

Discussion

Principal Results

In this study, we developed a novel tool to evaluate the quality of health information in online newspaper articles by reviewing previous studies and existing tools. The HIQUAL consists of 5 domains, namely reliability, usefulness, ease of understanding, sufficiency, and transparency. We found the HIQUAL to have high interrater agreement. After evaluating a total of 16 online newspaper articles, the HIQUAL was highly correlated with two other tools—DISCERN and QUEST. The results of the analysis, divided into treatment articles and diagnosis and prevention articles, also showed similar results to the overall analysis.

Comparison With Prior Work

In the process of developing the HIQUAL, a variety of previously developed quality assessment tools were compared. Among them was the DISCERN instrument, which consists of 16 questions and assesses the quality of treatment information for diseases; several studies have demonstrated its validity and reliability [26,46]. In Korea, Park et al used the translated version of the tool to evaluate the quality of health information websites that provide information on diseases such as breast cancer, asthma, depression, and obesity [47]. In addition, DISCERN was also used to evaluate websites that provide information on colorectal cancer [48], hepatitis B [49], and precocious puberty [50]. The DISCERN instrument has the advantage of being useful both to experts and the general public for conducting systematic comprehensive assessments, but it has not yet been validated in Korea and may be difficult to apply to information other than that relating to diseases and treatment [19,47]. The HONcode consists of 8 ethical codes to follow when providing information and has been used in Korea for evaluating online medical information on diabetes and thyroid cancer [51,52]. The code of ethics includes information delivery entities, sources of information, and justification items; however, the limitation is that these items do not guarantee the accuracy of the content. MedCERTAIN is part of an international project for the safe use of the internet, which requires health information providers to comply with its standards and assess compliance, based on items corresponding to standard metadata [53]. In Korea, it has been used to evaluate websites that provide information on dementia [54]. The HONcode and MedCERTAIN are better suited for evaluating platforms or websites that provide information, rather than evaluating individual online newspaper articles. QUEST was recently developed for evaluating online health information and has proven to be comparable to DISCERN [6]. This tool uses the 6 criteria of authorship, attribution, conflict of interest, currency, complementarity, and tone. However, the questions related to usefulness and understandability in HIQUAL's criteria were not used in QUEST. In addition, QUEST uses indirect evaluations on the basis of the tone to assess for exaggeration or error, which may facilitate more objective evaluations by nonexpert evaluators but may also lead to a somewhat less accurate assessment.

Since the number of consumers using internet health information has increased in Korea, Son presented criteria for quality evaluation based on prior studies that reviewed domestic and foreign health information websites [11]. This study faced limitations in applying these criteria because the actual assessment was not carried out. Kim et al developed user-specific (professional, operator, and general public) assessment tools for internet health information and have confirmed the reliability of these tools with the public and experts [12]. However, it is difficult to apply the tool to other types of media because it is intended to evaluate websites only. The tool developed by the Korean Medical Association consists of 14 questions across 5 categories: whether the information was reliable (reliability), whether it was helpful to readers (usefulness), whether the readers understood the contents

XSL•FO
RenderX

(understandability), whether the information was provided without omission (completeness), and whether this health information was organized for public interest (publicity). Although experts were asked to evaluate the suitability of each item, the development process of the evaluation items, validity, and reliability were not demonstrated [13]. The Korean Medical Association also evaluated the health information of newspaper articles and the internet through its own standards to identify the health-related information sought by consumers. There was a limit to the representativeness of the subjects who conducted the evaluation based on the proposed criteria.

A variety of tools have been used, but most of them are limited in the purpose and objective of evaluation and may not be suitable for evaluating online newspaper articles. Newspaper articles cover a wide range of content, ranging from diagnosing, treating, and preventing diseases to health care and new scientific discoveries. It should be considered that insufficient information, as well as information supported by scant evidence, can be communicated to an unspecified number of people in this process. Additionally, it is also necessary to convey sufficient information from an objective and independent perspective, as well as an appropriate understanding of the uncertainty of scientific research [17]. However, according to a study in the United States that analyzed online newspaper articles on drugs, many articles did not provide sufficient information, including the side effects and the cost of the medications [20]. There are nonprofit websites such as HealthNewsReview.org that have created their own criteria to address these problems, but the evaluations by this site are currently suspended. An analysis of articles on the new guidelines for diagnosing high blood pressure using those criteria, released in 2017, showed that only 33 of the 100 articles mentioned the benefits and risks that could arise from the changed guidelines, while only 2 articles mentioned conflicts of interest [55]. The validity and reliability of the criteria were not identified, but as these criteria also target online articles, they were considered in the development process of this study.

Limitations

There are points to consider when applying the HIQUAL. First, it is aimed at health-related online newspaper articles, so it is difficult to apply it to other forms of media or information. A tool that can cover a variety of evaluation targets may degrade its own accuracy and value, and well-made existing tools such as DISCERN and HONcode also limit their evaluation targets. Consequently, a tool developed in accordance with the characteristics and purpose of the evaluation target can be determined well in advance to evaluate the evaluation medium. Furthermore, it can be used more effectively when applied to articles that require neutral and sufficient information delivery, such as new medical findings or treatments, than to articles that convey well-known universal knowledge. With the evolution of technology over time, the issues of inaccurate online information will continue to arise, so our tool can be useful in targeting online newspaper readers.

The second limitation is the problem of the users. The result of the interrater agreement of the third article, which had a lower level of agreement than the other articles, was analyzed. We confirmed that the level of agreement increased when we examined the results of 4 medical specialists. To use this tool properly, judgment on medical validity or errors is necessary, and considering this, it is desirable for medical doctors or professionals in the field of health care to participate in the evaluation. In addition, as the results of the interrater agreement show, there may be some differences in the results of the assessment among raters. Therefore, in such a case, quality evaluation may be considered by two or more evaluators individually, and a final evaluation may be derived through consultation. During the actual application, it may also be useful to train the raters in advance to fully understand the assessment tools or to organize and operate an evaluation group where raters who are familiar with the tools can continue to participate.

Third, although the criteria and tools have been revised and identified by repeating the process of verifying reliability, the fact that the confirmation of the reliability and validity of the final completed tool was made using a relatively small number of articles may constitute another limitation. The fact that the final evaluation was made by a small number of experts and that they were not a representative group could also be a limitation of this study.

Conclusions

This study developed a new evaluation tool, the HIQUAL, for performing quality assessment of health-related online newspaper articles. For more effective use of the tool, it is desirable to establish a system that continuously monitors and evaluates health-related articles and delivers evaluation results to consumers so that they can make accurate judgments. Moreover, this tool could help information producers, such as journalists or reporters, produce quality health-related information articles. With the quality assessment tool, data producers can provide accurate and understandable information for online health-related articles. Using the HIQUAL, it will be possible to establish a platform that conducts continuous evaluations and regularly publishes the results, giving audiences access to high-quality health-related online newspaper articles. In addition, this tool will guide practitioners in the medical field in advancing sound strategies for disseminating health information among the general public and promote collaboration between experienced medical practitioners and news sites. Against the backdrop of the increasing number of health-related newspaper articles, as well as concerns about their quality and accuracy, this tool may be useful for assessing the quality of online health information in Korea.

Acknowledgments

This study was supported by the Korean Medical Association research project (grant 2019-01).

XSL•FO
RenderX

Authors' Contributions

All authors contributed to conceptualization, data curation, and evaluation. NL, SWO, and BC drafted the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality assessment tool for health-related newspaper article (Korean version).
[PDF File (Adobe PDF File), 85 KB - jmir_v23i7e24436_app1.pdf]

References

1. Kim S, Oh J, Lee Y. Health Literacy: An Evolutionary Concept Analysis. The Journal of Korean Academic Society of Nursing Education 2013 Nov 30;19(4):558-570. [doi: 10.5977/jkasne.2013.19.4.558]
2. Weiner SS, Horbacewicz J, Rasberry L, Bensinger-Brody Y. Improving the Quality of Consumer Health Information on Wikipedia: Case Series. J Med Internet Res 2019 Mar 18;21(3):e12450 [FREE Full text] [doi: 10.2196/12450] [Medline: 30882357]
3. D'Cruz M, Kini R. The effect of information asymmetry on consumer driven health plans. In: Integration and Innovation Orient to E-Society Volume 1. Boston, MA: Springer; 2007:353-362.
4. Kim MC, Lee DC. Influence and extending effect on healthcare service by information communication technology: focused on internet technology for dissolution planning of information inequality between physician and patient. Journal of Electronic Commerce Research 2015;5:29-46.
5. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the Accuracy and Readability of Online Health Information for Patients With Pancreatic Cancer. JAMA Surg 2016 Sep 01;151(9):831-837. [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]
6. Robillard JM, Jun JH, Lai J, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
7. Oh YS, Cho Y. Exploring the Limitations in the Use of Online Health Information and Future Direction: Focused on Analysis of Expert Knowledge in the Frame of Ignorance. Health and Social Welfare Review 2019 Jun;39(2):358-393. [doi: 10.15709/hswr.2019.39.2.358]
8. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
9. HONcode. 2020. URL: http://www.hon.ch/HONcode [accessed 2020-01-04]
10. Eysenbach G, Yihune G, Lampe K, Cross P, Brickley D. MedCERTAIN: quality management, certification and rating of health information on the Net. Proc AMIA Symp 2000:230 [FREE Full text] [Medline: 11079879]
11. Son A. Criteria for evaluating health information sites on the internet: cross sectional survey. J KOSHIS 2000;2(1):73-89.
12. Kim MJ, Kang NM, Kim SW, Rhyu SW, Chang H, Hong SK, et al. Development of an Evaluation Checklist for Internet Health/Disease Information. J Korean Soc Med Inform 2006;12(4):283. [doi: 10.4258/jksmi.2006.12.4.283]
13. Korean Academy of Medical Sciences. URL: https://www.kams.or.kr/business/judge/sub2/ [accessed 2020-03-14]
14. Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW. Making Quality Health Websites a National Public Health Priority: Toward Quality Standards. J Med Internet Res 2016 Aug 02;18(8):e211 [FREE Full text] [doi: 10.2196/jmir.5999] [Medline: 27485512]
15. Charnock D. The DISCERN handbook. Quality criteria for consumer health information on treatment choices. Radcliffe: University of Oxford and the British Library; 1998.
16. Mitchell A. Americans Still Prefer Watching to Reading the News – and Mostly Still Through Television. URL: https://www.journalism.org/2018/12/03/americans-still-prefer-watching-to-reading-the-news-and-mostly-still-through-television/ [accessed 2020-04-16]
17. Newspapers Deliver Across the Ages. URL: https://www.nielsen.com/us/en/insights/article/2016/newspapers-deliver-across-the-ages/ [accessed 2020-03-26]
18. Korea Press Foundation. URL: https://www.kpf.or.kr/front/user/main.do [accessed 2020-03-26]
19. Jeong SH, Kim J, Kim T, Park S, Shin Y, Lee S. Survey on the consumer preference for the internet health information of the patients' online community members. J Korean Soc Med Inform 2007;13(3):207. [doi: 10.4258/jksmi.2007.13.3.207]
20. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. N Engl J Med 2000 Jun 01;342(22):1645-1650. [doi: 10.1056/NEJM200006013422206] [Medline: 10833211]

XSL•FO
RenderX

21. Roh S, Yoon Y. Analyzing online news media coverage of depression. Korean Journal of Communication and Information 2013;61:5-27.
22. Kim MK, Lee YS, Cho SY. Korean newspaper coverage of infertility (1962-2013). Korean Society for Journalism and Communication Studies 2014;58(6):329-361.
23. Nastasi A, Bryant T, Canner JK, Dredze M, Camp MS, Nagarajan N. Breast Cancer Screening and Social Media: a Content Analysis of Evidence Use and Guideline Opinions on Twitter. J Cancer Educ 2018 Jun;33(3):695-702. [doi: 10.1007/s13187-017-1168-9] [Medline: 28097527]
24. HealthNewsReview. URL: https://www.healthnewsreview.org/ [accessed 2020-11-01]
25. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 2015 Dec 11;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
26. Ademiluyi G, Rees CE, Sheard CE. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. Patient Educ Couns 2003 Jun;50(2):151-155. [doi: 10.1016/s0738-3991(02)00124-6] [Medline: 12781930]
27. Winker MA, Flanagin A, Chi-Lum B, White J, Andrews K, Kennett RL, et al. Guidelines for medical and health information sites on the internet: principles governing AMA web sites. JAMA 2000;283(12):1600-1606. [doi: 10.1001/jama.283.12.1600] [Medline: 10735398]
28. Lee JW, Yoon SC, Lee S. Hierarchical Clustering Analysis of Information Quality Dimensions. Entrue Journal of Information Technology 2003;2(2):91-99.
29. Seo MK, Chung YC, Oh YK, Lee SK. Analysis of internet health information. Sejong City, Korea: Korea Institute for Health and Social Affairs; Mar 2000:68-76.
30. NAVER. URL: http://www.naver.com/ [accessed 2020-11-01]
31. Shin S. Health information Lack of vitamin D after menopause increases the risk of hypertension and hyperglycemia metabolic syndrome. URL: http://news.heraldcorp.com/view.php?ud=20180403000168 [accessed 2019-02-05]
32. Han SH. Smoking also worsens hearing... 1.7 times lower than non-smokers. URL: https://www.sedaily.com/NewsView/1RY41VC6QR [accessed 2019-02-05]
33. Development of blood test to detect Alzheimer's early. Kang MK. URL: https://medicalreport.kr/news/view/46895 [accessed 2019-02-05]
34. Kong KA. Statistical Methods: Reliability Assessment and Method Comparison. Ewha Med J 2017;40(1):9-16. [doi: 10.12771/emj.2017.40.1.9]
35. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971;76(5):378-382. [doi: 10.1037/h0031619]
36. Kim MS, Song KJ, Nam CM, Jung IK. A Study on Comparison of Generalized Kappa Statistics in Agreement Analysis. Korean Journal of Applied Statistics 2012 Oct 31;25(5):719-731. [doi: 10.5351/kjas.2012.25.5.719]
37. Gwet K. Benchmarking inter-rater reliability coefficients. In: Handbook of inter-rater reliability 3rd edn. Gaithersburg, MD: Advanced Analytics, LLC; 2012:121-128.
38. Ko MM, Lee JA, Kang B, Park T, Lee J, Lee MS. Interobserver reliability of tongue diagnosis using traditional korean medicine for stroke patients. Evid Based Complement Alternat Med 2012;2012:209345 [FREE Full text] [doi: 10.1155/2012/209345] [Medline: 22474492]
39. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013 Apr 29;13:61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
40. Kendall MG. A new measure of rank correlation. Biometrika 1938 Jun 01;30(1-2):81-93. [doi: 10.1093/biomet/30.1-2.81]
41. Hauke J, Kossowski T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quaestiones Geographicae 2011 Jun 24;30(2):87-93. [doi: 10.2478/v10117-011-0021-1]
42. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71 [FREE Full text] [Medline: 23638278]
43. Western-style diet and obesity changed to 'Korean colon cancer topography'. The Kukmin Daily. URL: http://news.kmib.co.kr/article/view.asp?arcid=0012725640&code=61121911&cp=nv [accessed 2019-03-25]
44. Lee SY. Now obesity is also controlled by genes. EDaily. URL: https://www.edaily.co.kr/news/read?newsId=01679366619339464&mediaCodeNo=257&OutLnkChk=Y [accessed 2019-03-25]
45. Lim YJ. When high blood pressure standards are strengthened, there is less work to be caught in the 'back neck'. Seoul Daily. URL: https://www.sedaily.com/NewsView/1S5R9TG202 [accessed 2019-03-28]
46. Rees CE, Ford JE, Sheard CE. Evaluating the reliability of DISCERN: a tool for assessing the quality of written patient information on treatment choices. Patient Educ Couns 2002 Jul;47(3):273-275. [doi: 10.1016/s0738-3991(01)00225-7] [Medline: 12088606]
47. Park JH, Cho BL, Kim YI, Shin YS, Kim Y. Assessing the Quality of Internet Health Information Using DISCERN. J Korean Soc Med Inform 2005;11(3):235. [doi: 10.4258/jksmi.2005.11.3.235]
48. Sohn DK, Choi HS, Lee DU, Lee SJ, Lee JS, Lee YS. The Quality Evaluation of Information of Websites on Colorectal Cancer Using the DISCERN Instrument. Journal of the Korean Society of Coloproctology 2005;21(4):247-254.

XSL•FO
RenderX

49. Jung YG, Ahn DS, Choi YJ. Evaluation of hepatitis B medical information quality using DISCERN. The Journal of the Institute of Internet, Broadcasting and Communication 2010;10(5):63-67.
50. Kim BK, Park SH, Sung HU, Park SY, Kwon YS, Jun YH, et al. Evaluation of Information of Websites on Precocious Puberty. Ann Pediatr Endocrinol Metab 2012;17(1):27. [doi: 10.6065/apem.2012.17.1.27]
51. Min YJ, Oh MY, Jung YG. Reliability assessment of diabetes care information using HONCODE using embedded Linux system. 2012 Presented at: The Institute of Electronics and Information Engineers Conference; 2012; Jeongseon p. 929-931.
52. Heo J, Jung Y, Sihn S, Kim J. Evaluation of thyroid cancer medical information sites using HONCODE. Journal of Service Research and Studies 2013;3(2):45-52.
53. Jung YC. The quality improvement method of health information on internet-introduction for MedCERTAIN project. Health and Welfare Policy Forum 2001;10:84-94.
54. Heo J, Park C, Jung Y. Quality assessment of dementia medical information using MedCERTAIN. 2012 Presented at: The Institute of Electronics and Information Engineers Conference; 2012; Jeongseon p. 951-953.
55. Moynihan RN, Clark J, Albarqouni L. Media Coverage of the Benefits and Harms of the 2017 Expanded Definition of High Blood Pressure. JAMA Intern Med 2019 Feb 01;179(2):272-273 [FREE Full text] [doi: 10.1001/jamainternmed.2018.6201] [Medline: 30592488]

Abbreviations

AC: agreement coefficient
DQ: data quality
HIQUAL: Health Information Quality Assessment Tool
HONcode: Health on the Net Foundation Code of Conduct
QUEST: Quality Evaluation Scoring Tool

Edited by R Kukafka; submitted 21.09.20; peer-reviewed by S Ghaddar, R Haase; comments to author 18.11.20; revised version received 11.01.21; accepted 14.06.21; published 29.07.21.

Please cite as:
Lee N, Oh SW, Cho B, Myung SK, Hwang SS, Yoon GH
A Health Information Quality Assessment Tool for Korean Online Newspaper Articles: Development Study
J Med Internet Res 2021;23(7):e24436
URL: https://www.jmir.org/2021/7/e24436
doi:10.2196/24436
PMID:34326038

©Naae Lee, Seung-Won Oh, Belong Cho, Seung-Kwon Myung, Seung-Sik Hwang, Goo Hyeon Yoon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Content Analysis and Quality Evaluation of Cesarean Delivery–Related Videos on YouTube: Cross-sectional Study

Kyong-No Lee[1], MD; Yeon Ji Joo[1], MD; So Yeon Choi[1], MD; Sung Taek Park[1,2], MD, Prof Dr; Keun-Young Lee[1], MD, Prof Dr; Youngmi Kim[2], PhD; Ga-Hyun Son[1,2], MD, Prof Dr

[1]Department of Obstetrics and Gynecology, Hallym University Kangnam Sacred Heart Hospital, Seoul, Republic of Korea

[2]Institute of New Frontier Research, College of Medicine, Hallym University, Chuncheon, Republic of Korea

Corresponding Author:
Ga-Hyun Son, MD, Prof Dr
Department of Obstetrics and Gynecology
Hallym University Kangnam Sacred Heart Hospital
Daelim-Dong, Yeoungdeungpo-Gu 948-1
Seoul, 07441
Republic of Korea
Phone: 82 2 829 5114
Fax: 82 2 849 4468
Email: ntr5017@naver.com

Abstract

Background: YouTube is one of the most popular open-access video-sharing websites, and it is also used to obtain health care information. Cesarean delivery is the most common major surgical intervention in many countries. Videos related to cesarean delivery have also been uploaded to YouTube. However, no study has explored the overall quality of cesarean delivery videos on the platform.

Objective: The objective of this study was to analyze the content and evaluate the quality of the most frequently viewed videos related to cesarean delivery that are accessible on YouTube.

Methods: We searched for a total of 18 terms by combining the 6 terms retrieved from Google AdWords and the 3 terms *c section*, *cesarean section*, and *cesarean delivery*, which are used interchangeably. Videos were sorted by view count, and the 100 videos with the highest view counts were chosen. The number of views, duration, likes and dislikes, content type, and source of each video were recorded. In evaluating the quality of the videos, we referred to a previous study. Additionally, we developed a detailed scoring method that comprehensively evaluates the videos related to cesarean delivery by including the necessary information for each element of the cesarean delivery and whether scientific evidence was presented.

Results: Of the 100 videos analyzed, the most prevalent content (n=28) was videos that contained the actual surgical procedure of a cesarean delivery, and the most common source of cesarean delivery videos was physicians (n=30). Videos directly related to cesarean delivery, such as explanation of the surgery and the actual surgical procedure, were mainly uploaded by medical groups and scored higher than the videos indirectly related to cesarean delivery, which were mainly uploaded by nonmedical groups. In addition, videos directly related to cesarean delivery were more often uploaded earlier in time, with lower like ratios compared to indirect videos.

Conclusions: YouTube is currently not an appropriate source for patients seeking information on cesarean delivery.

(J Med Internet Res 2021;23(7):e24994) doi:10.2196/24994

KEYWORDS

cesarean delivery; YouTube; internet; quality of information

Introduction

Cesarean delivery is the most common surgical intervention in many countries [1]. Based on data from 169 countries, 29.7 million births worldwide were estimated to be by cesarean delivery in 2015, which was almost double the number of births in 2000 (16 million) [2]. Cesarean delivery can be lifesaving for the fetus, mother, or both in certain cases, such as dystocia, placenta previa, or abnormal fetal presentation; however, the rapid increase in the rate of cesarean births without any evidence

XSL•FO
RenderX

of associated reduction in maternal or neonatal morbidity or mortality raises concerns that cesarean delivery is being performed for the convenience of the patients or physicians even when it is not required [3]. In recent times, the rate of cesarean delivery due to maternal requests has increased to 8% due to fear of labor pain, anxiety about fetal injury, urinary incontinence, or pelvic floor dysfunction [4]. The fear of litigation among physicians has also played a role in the increase in cesarean delivery rates. Moreover, the autonomy of the patient tends to be a more important consideration in deciding the method of delivery. Therefore, it is important for patients to obtain accurate information about cesarean delivery based on scientific evidence.

In the last decade, social media has emerged as an important source of health care–related information. Altogether, 80% of adults in the United States have used the internet to access health care information [5,6]. Among the web-based resources, YouTube, an open-access video-sharing website, is among the three most popular websites, with more than 4 billion videos viewed daily and more than 500 hours of video content uploaded every minute [7]. YouTube is becoming an increasingly popular platform for users to obtain, share, and discuss health information. In providing information, the social media format has the advantage of possible timely updates; however, social media platforms may contain misleading and inappropriate information because there is a lack of regulation of the content and no peer review process [8-10]. To date, no study has yet evaluated cesarean delivery–related information on YouTube. Therefore, the purpose of this study was to describe and analyze the content of the most-viewed videos of cesarean delivery on YouTube to identify features of cesarean delivery–related videos that were watched by the general public. We also evaluated the quality of the videos related to cesarean delivery on YouTube to determine whether accurate and important information was being delivered.

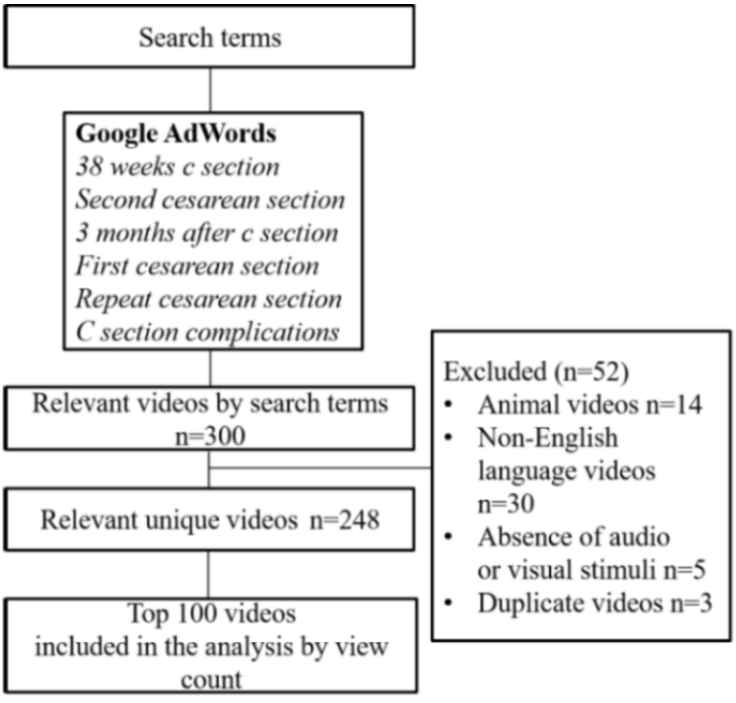
Methods

Retrieval of Cesarean Delivery–Related YouTube Videos

We intended to include representative videos about cesarean delivery that the public could access. The Google keyword search tool Google AdWords [11], a method used by Williams et al [12], was used to identify appropriate search terms that the public would use to explore the term “cesarean section” on YouTube. From the original keyword *cesarean section*, a list of popular terms, such as *38 weeks c section*, *second c section*, *3 months after c section*, *first c section*, *repeat c section*, and *c section complications*, was retrieved. Each term was queried by combining 3 terms, namely, *c section*, *cesarean section*, and *cesarean delivery*, and a total of 18 terms were queried. A YouTube search was conducted on April 2, 2021. As the goal of the search was to identify the videos that the public were watching most frequently, the videos were sorted by view count using the YouTube advanced search options. The criteria for including the videos were as follows: (1) English language used, (2) primary content related to cesarean delivery, and (3) acceptable audiovisual quality. The exclusion criteria were as follows: (1) languages other than English, (2) poor audio or visual clarity, (3) animal videos, and (4) duplicate videos.

For the search terms, the top 300 initial videos were included for review, as determined by the Relevance filter according to YouTube’s algorithm. A total of 52 videos were excluded (14 animal videos, 30 non–English-language videos, 5 videos with poor audio or visual clarity, and 3 duplicate videos). A list of the top 100 videos was populated based on view count, and this list served as the basis for the subsequent analysis. A description of the search strategies is presented in Figure 1.

Figure 1. Methodology of selection of YouTube videos for analysis.



XSL•FO
RenderX

Video Assessment

For each video, we collected the objective (video title, number of views, uploaded day, uploader's name, length of the video, days since upload, total number of "likes" and "dislikes" as depicted by the "thumbs up" and "thumbs down" icons, details and sources related to the cesarean delivery included in the video) and subjective data (the purpose and type of content).

Based on the content in the videos, the videos were categorized into five groups: (1) explanations of surgery (providing general cesarean delivery–related information), (2) surgery procedure (showing or explaining detailed surgical procedure techniques and processes), (3) personal experiences (sharing personal experiences and feelings related to pregnancy and delivery), (4) postpartum care (providing information on postsurgical care, eg, nutrition, exercise, and wound care), and (5) others (mock practice videos in various circumstances, appreciation of medical dramas related to cesarean delivery, and description of surgical instruments). We classified videos that include explanations of surgery and surgical procedures as videos directly related to cesarean delivery, and videos that included personal experiences or postpartum care were classified as videos indirectly related to cesarean delivery (Table 1).

Table 1. Characteristics of videos related to cesarean delivery on YouTube (N=100).

Variable	Description	Value, n
Content		
Directly related		
Explanations of surgery	Provide general cesarean delivery–related information	19
Surgery procedure	Show or explain detailed surgical procedure techniques and processes	28
Indirectly related		
Personal experiences	Share personal experiences and feelings related to pregnancy and delivery	19
Postpartum care	Provide information about postsurgical care (eg, nutrition and exercise, wound care)	24
Others	Mock practice video in various circumstances, appreciation of medical dramas related to cesarean delivery, and description of surgical instruments	10
Source		
Medical		
Academic	Authors are affiliated with a university	12
Physician	Authors are not affiliated with a university but are physicians	30
Nonmedical		
Patient	Woman who has already delivered or is currently pregnant, or her husband	24
Commercial	Attention to a product or service	8
Paramedical	Allied health therapist, physiotherapist, or dietitian	26

Based on their authorship, the videos were classified into five basic groups: **(**1) academics (authors were affiliated with a university), (2) physicians (authors were not affiliated with a university but were physicians), (3) patients (a woman who had already delivered or was currently pregnant, or her husband), (4) commercial establishments (attention to a product or service), and (5) paramedical (allied health therapist, physiotherapist, or dietitian). We further categorized the videos uploaded by academic and physician groups into the medical group, and those uploaded by patients, commercial establishments, and paramedical groups were categorized into the nonmedical group (Table 1).

Quality Assessment

Because there are no established standards for evaluating video quality, we prepared an arbitrary scoring system by referring to a previous study [13-15]. The evaluation factors were divided into the part that evaluates the general quality of the video, whether important information on cesarean delivery was included and explained, and how much scientific evidence was specified (Textbox 1). For general video quality and flow of video contents, each parameter was scored on a scale of 1 to 3. The information on cesarean delivery was divided into 5 elements and scored as follows depending on the degree of explanation: 0 points, not mentioned; 1 point, mentioned briefly; 2 points, mentioned in detail. For videos based on scientific evidence, there were two subdivided items: 0 points were given if there was no mention, and 1 point was given. Thus, the total score of the 5 items ranged from a minimum of 2 to a maximum of 18 points. Three professional obstetricians (two professors at the University Hospital of Obstetrics and one obstetrician fellow) independently evaluated the quality of each video, and the score used for the analysis was the average of the 3 scores.

To assess the popularity of the videos, we used the like ratio (like × 100/[like + dislike]), view ratio (number of views/days), and video power index (VPI) (like ratio × view ratio/100).

XSL•FO
RenderX

Textbox 1. Predetermined list of evaluation factors for the quality of videos on YouTube related to cesarean delivery.

General quality (Poor: 1 point; moderate: 2 points; good: 3 points)

- Overall video quality (audio and video)
- Flow of contents in videos

Degree to which information is helpful to viewers (Not mentioned: 0 points; mentioned briefly: 1 point; mentioned in detail: 2 points)

- Indication of cesarean delivery
- Maternal or fetal complications
- Surgical process
- Preoperative preparation, anesthesia
- Postoperative management, postpartum care

Scientific evidence (No: 0 points; yes: 1 point)

- Clearly states sources of information
- Provides details of where to obtain additional information on the video topic

Statistical Analysis

Data are shown as median (range) for continuous variables and as n (%) for categorical variables. Comparisons were made between videos directly related to cesarean delivery and videos with indirectly related contents and between the medical and nonmedical groups using the Mann-Whitney *U* test. Comparisons of the difference in uploaded contents between medical and nonmedical groups were analyzed by the Fisher exact test. The Kruskal-Wallis test with Bonferroni corrections was used to compare scores according to the contents of the uploaded videos. Statistical analysis was conducted using SPSS software, version 27.0 (IBM Corporation). Statistical significance was set at $P<.05$. The reliability between the YouTube videos and the scores by the three obstetricians on the criteria for the items was assessed using intraclass correlation coefficients.

Ethics Statement

Institutional review board approval was waived for this study because only publicly available data were used.

Results

The top 100 videos related to cesarean delivery had a collected total of 285,666,251 views (median 1,146,376, range 253,267-31,326,580). The descriptive features of the cesarean delivery–related videos on YouTube are shown in Table 2. The highest view count for a video was 31,326,580. This video was uploaded by a medical media channel in 2008, and it contains general information on cesarean delivery such as operation indication and pelvic anatomy; it also explains detailed surgical processes, preoperative preparation, and postoperative management. This video also received the maximum number of likes (77,164).

Table 2. Descriptive features of videos related to cesarean delivery on YouTube (N=100).

Variable	Median (range)
Views, n	1,146,376 (253,267-31,326,580)
Video length (minutes)	5.83 (0.25-34.82)
Time on YouTube (days)	1538.5 (39-5044)
Comments, n	150 (0-5127)
Likes (thumbs up), n	5086 (65-77,164)
Dislikes (thumbs down), n	570 (11-13,534)
Like ratio	88 (60-99)
View ratio	784 (101-73,366)
Video power index	706 (95-66,857)

The median length was 5.83 minutes (range 0.25-34.82), and the majority of videos (80/100, 80%) did not exceed 12 minutes. Videos were uploaded to YouTube approximately 1538.5 days previously on median (range 39-5044), and the most videos were uploaded in 2019. Because the videos included in this study were sorted by number of views, considering that the accumulated views of the recently uploaded videos could be fewer than those uploaded earlier, we observed that the number of uploaded videos began to increase rapidly from 2012 (Table 3).

XSL•FO
RenderX

Table 3. Number of videos included in the study (N=100) by year of upload.

Year	Uploaded videos, n
2007	2
2008	1
2009	0
2010	2
2011	2
2012	5
2013	8
2014	9
2015	12
2016	9
2017	14
2018	13
2019	16
2020	6
2021	1

Table 1 shows a description of the categorization according to video contents and authorship. The most prevalent content (n=28) was videos that contained the actual surgical procedure of a cesarean delivery, and in many cases (20/28, 71%), the videos were uploaded by physicians. The second most commonly uploaded videos were videos with information on postpartum care, including postoperative exercises to achieve recovery, nutritional care, and wound care. When the video content was categorized as directly related to cesarean delivery or indirectly related according to the content characteristics of the video, there were 47 directly related videos containing explanations on surgery and actual surgery procedures and 53 indirectly related videos. The source that uploaded the most videos was physicians. When academics and physicians were collectively referred to as a medical group, 42 videos were uploaded by the medical group and 58 videos by the nonmedical group. When analyzing the relationship between the source of the videos and the content, the medical group mainly uploaded videos directly related to cesarean delivery, such as explanations or detailed surgical procedures for cesarean delivery (37/42, 88%), while the nonmedical group mainly uploaded videos about personal experiences and postoperative care (40/58, 69%) (P<.001).

To evaluate whether the videos related to cesarean delivery contained accurate and important information on cesarean delivery and whether scientific evidence was presented, we created a detailed scoring method (Textbox 1). The median score was 6 (range 1-16); the video with the highest score was a video containing a well-organized general description of cesarean delivery with the most views, and the video with the lowest score was mainly composed of personal pregnancy photos without including contents related to cesarean delivery. When evaluating the score by each type of content, videos containing personal experience (median score 4, range 2-6) scored significantly lower than videos containing other contents, such as explanations of surgery (median score 9, range 3-16); P<.001), surgery procedures (median score 6, range 1-12; P<.001) and postpartum care videos (median score 6, range 3-11; P=.001). There was a high degree of correlation between the reviewers (intraclass correlation coefficient 0.908, 95% CI 0.872-0.935; P<.001).

Next, we created and analyzed the like ratio, view ratio, and VPI to evaluate which videos people were interested in and liked. The video with the highest VPI was a video about the childbirth of an Indian actress, posted by Bollywood Trends, which focuses on birth news rather than cesarean delivery–related information.

We further analyzed how videos related to cesarean delivery uploaded on YouTube differ according to the uploaded content, source, and time when the video was uploaded to YouTube. We compared videos uploaded by the end of 2015 and videos uploaded after 2015, which was the midpoint between 2007 and 2021, and we also compared the videos according to the contents and source.

When we compared videos directly related to cesarean delivery with those indirectly related to cesarean delivery according to the content characteristics of the videos, the videos directly related to cesarean delivery were uploaded earlier (direct vs indirect: median time on YouTube 2247 days, range 266-5044, vs 1298 days, range 39-3821;P=.02), with a lower like ratio (median 85, range 62-98, vs median 91, range 60-99; P=.003) and higher score (median 7, range 1-16, vs median 5, range 2-11; P<.001) compared to indirect videos (Table 4). When analyzing the proportion of videos directly related to cesarean delivery compared to all videos, 27/41 videos (66%) directly related to cesarean delivery were uploaded by the end of 2015, but this number decreased to 20/59 (34%) after 2015 (P=.009).

XSL•FO
RenderX

Table 4. Comparison of the content of the videos (N=100).

Variable	Value, median (range)		*P* value
	Directly related (n=47)	Indirectly related (n=53)	
Views, n	1,616,358 (311,534-31,326,580)	960,159 (253,267-17,553,197)	.23
Video length (minutes)	5.62 (0.25-32.35)	5.83 (0.95-34.82)	<.99
Time on YouTube (days)	2247 (266-5044)	1298 (39-3821)	.02
Comments, n	242 (0-5127)	149 (0-3586)	.78
Likes (thumbs up), n	4514 (65-77,164)	5217 (195-58,550)	.60
Dislikes (thumbs down), n	762 (14-13,534)	434 (11-12,521)	.27
Like ratio	85 (62-98)	91 (60-99)	.003
View ratio	768 (124-19,265)	786 (101-73,366)	<.99
Video power index	633 (116-18,128)	730 (95-66,857)	.60
Score	7 (1-16)	5 (2-11)	<.001

When analyzed according to source (Table 5), the median number of views and degree of popularity of the videos represented by VPI did not significantly differ between the videos uploaded by the medical group and those by the nonmedical group. However, videos uploaded by the medical group showed significantly higher scores than those by the nonmedical group (median 8, range 4-16, vs median 5, range 1-11; *P*<.001).

When analyzed according to the date when the video was uploaded (Table 6), the videos uploaded after 2015 received more comments (median 271, range 0-5127, vs median 82, range 0-3944; *P*=.005) and had a higher VPI than videos uploaded by the end of 2015 (median 858.5, range 227-66,857, vs 491, range 95-7925; *P*=.005).

Table 5. Comparison according to the source of the videos (N=100).

Variable	Value, median (range)		*P* value
	Medical group (n=42)	Nonmedical group (n=58)	
Views, n	1,222,121.5 (281,480-31,326,580)	1,124,691 (253,267-5,286,769)	.84
Video length (minutes)	8.19 (1.42-32.1)	5.38 (0.25-34.82)	.37
Time on YouTube (days)	1976.5 (266-5,044)	1494.5 (39-3821)	.31
Comments, n	256 (0-3944)	148 (0-5127)	.53
Likes (thumbs up), n	4658.5 (65-77,164)	5087 (195-72,024)	.90
Dislikes (thumbs down), n	571 (14-8445)	467 (11-13,534)	.90
Like ratio	86 (62-99)	91 (60-99)	.07
View ratio	717.5 (124-7949)	795 (101-73,366)	.54
Video power index	614.5 (116-7796)	740 (95-66,857)	.34
Score	8 (4-16)	5 (1-11)	<.001

XSL•FO
RenderX

Table 6. Comparison according to the time of upload of the videos to YouTube.

Variable	Value, median (range)		*P* value
	By the end of 2015 (n=41)	After 2015 (n=59)	
Views, n	1,589,552 (253,267-31,326,580)	998,717 (281,480-12,445,056)	.42
Video length (minutes)	5.62 (0.25-27.32)	7.82 (0.95-34.82)	.68
Comments, n	82 (0-3944)	271 (0-5127)	.005
Likes (thumbs up), n	3213 (65-77,164)	6189 (163-72,024)	.049
Dislikes (thumbs down), n	574 (11-13,534)	488 (14-4518)	.62
Like ratio	84 (61-98)	91 (60-99)	.004
View ratio	610 (101-11,254)	1040 (274-73,366)	.004
Video power index	491 (95-7925)	858.5 (227-66,857)	.005
Score	6 (1-16)	6 (2-13)	.55

Discussion

Principal Findings

Our study identified that the most viewed video about cesarean delivery was a video that was uploaded in 2008 by Nucleus Medical Media, which is a company that specializes in producing medical illustrations and animations. This video is a well-organized video containing overall information on cesarean delivery, such as operation indication, pelvic anatomy expressed by animations, and detailed step-by-step explanations of surgical procedures, preoperative preparation, and postoperative and postpartum management. This video also received the most likes and the highest score, and it also showed the fifth highest VPI. This video was uploaded in 2008, which is relatively early on YouTube, but it was still considered the best organized and most informative video about cesarean delivery. Of the videos included in this study, 41 videos were uploaded up by the end of 2015, and 59 videos had been uploaded since 2015. Our results showed that videos that were directly related to cesarean delivery were often uploaded at earlier dates, and the proportion of videos that were directly related to cesarean delivery out of all videos after 2015 decreased compared to that of videos uploaded by the end of 2015. In addition, although the difference was statistically insignificant, more than half of the videos from the medical group were uploaded by the end of 2015, while more videos were uploaded by the nonmedical group after 2015 (21/41, 51%, vs 21/59, 36%; *P*=.12). These results suggest that as YouTube becomes more popular and laypeople can easily access and produce content, the number of videos containing contents such as personal experiences and postpartum care being uploaded by laypeople is greater than the number of professional videos containing medical information on cesarean delivery being uploaded by medical groups. However, the videos uploaded after 2015 were more popular, as indicated by their VPIs and like ratios, and they also received more comments than the videos uploaded by the end of 2015. These results suggest that the quality and reliability of information provided by YouTube is not related to popularity. Also noteworthy is that although videos directly related to cesarean delivery had higher quality scores, their like ratios were notably lower than those of videos indirectly related to cesarean delivery; moreover, videos uploaded by medical groups scored higher, but their like ratios tended to be lower. The VPI, which is a comprehensive indicator reflecting popularity, did not show any differences between videos directly related to cesarean delivery and videos indirectly related to cesarean delivery or between videos uploaded by medical and nonmedical groups. These results showed that laypeople expressed their preferences regardless of the quality of the video. These results are similar to those of previous studies. Staunton et al [16] reviewed 50 videos regarding scoliosis and found that videos with greater educational quality were associated with a lower number of views. Ferhatoglu et al [17] recently reported an association between high VPI scores and low Sleeve Gastrectomy Scoring System scores in their review of sleeve gastrectomy videos on YouTube.

In addition, the results of this study showed that although the total possible score for each video was 18 points, the median score was 6 (range 1-16), which is relatively low. Among the videos, only 13 scored more than 10 points, and the remaining videos showed relatively low scores. The reason for this finding is that most of the videos received low scores in the evaluation items for the information elements related to cesarean delivery (median score 2, range 0-9). Moreover, there were few videos containing the indications for cesarean delivery and maternal or fetal complications that may occur after surgery (n=23 and n=21, respectively), which is important information to be aware of before undergoing a cesarean delivery. In addition, the majority of the videos (80/100, 80%) had a score of 0 in the scientific evidence category, showing that insufficient references were provided for the information in the video. It may not be appropriate to use the scoring method used in this study to evaluate the quality of videos on YouTube, where people can freely produce and upload videos on topics of interest. However, our results showed that videos on YouTube have limitations in providing general and well-organized scientific knowledge of cesarean delivery. It is possible that this limitation is the reason that while the number of searches for “cesarean section” on YouTube has been decreasing over time, the number of YouTube users and searches for “cesarean section” on the Google website has been increasing (Figure 2 and Figure 3). Although YouTube can be seen as a potentially useful medium to search for cesarean delivery–related knowledge and increase

XSL•FO
RenderX

awareness, the user must be aware that the information uploaded is not regulated and the quality of the content thus needs to be validated.

Figure 2. Search trend for the term *cesarean section* on YouTube.



Figure 3. Search trend for the term *cesarean section* on the Google website.



Limitations

This study had some limitations. First, although the assessment method used in this study was adapted from the DISCERN criteria [15], and it reflects the opinions of experts on cesarean delivery, the necessary information for cesarean delivery was included in the evaluation items; however, this method was created arbitrarily by us. Thus, more verification is needed to ensure that the assessment method is suitable for accurately evaluating the quality of videos on cesarean delivery. Second, we only analyzed videos that were in English; thus, sampling bias could have occurred. It is necessary to evaluate videos in other languages for a more comprehensive analysis of the features in videos on cesarean delivery.

Conclusion

This study is the first to analyze YouTube videos on cesarean delivery, and it contributes to a better understanding of the available information on cesarean delivery that is widely viewed on YouTube. Our results showed that the videos directly related to cesarean delivery, such as explanations of the surgery and actual surgical procedures, were mainly uploaded by medical groups and scored higher than the videos indirectly related to cesarean delivery, which were mainly uploaded by nonmedical groups. In addition, videos directly related to cesarean delivery were more often uploaded earlier in time, and the proportion of videos that were directly related to cesarean delivery decreased after 2015. In our results, when we used the scoring method to evaluate the accuracy of the important information on cesarean delivery, a majority of videos had low scores, showing that YouTube has limitations in delivering accurate information on cesarean delivery.

Acknowledgments

This research was supported by the Hallym University Research Fund 2021 (HURF-2021-05) and the Bio & Medical Technology Development Program of the National Research Foundation funded by the Korean government (Ministry of Science and Information and Communication Technology) (NRF-2019M3A9E8020538).

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Biccard BM, Madiba TE, Kluyts H, Munlemvo DM, Madzimbamuto FD, Basenero A, African Surgical Outcomes Study (ASOS) investigators. Perioperative patient outcomes in the African Surgical Outcomes Study: a 7-day prospective observational cohort study. Lancet 2018 Apr 21;391(10130):1589-1598. [doi: 10.1016/S0140-6736(18)30001-1] [Medline: 29306587]
2. Boerma T, Ronsmans C, Melesse DY, Barros AJD, Barros FC, Juan L, et al. Global epidemiology of use of and disparities in caesarean sections. Lancet 2018 Oct 13;392(10155):1341-1348. [doi: 10.1016/S0140-6736(18)31928-7] [Medline: 30322584]
3. Gregory KD, Jackson S, Korst L, Fridman M. Cesarean versus vaginal delivery: whose risks? Whose benefits? Am J Perinatol 2012 Jan;29(1):7-18. [doi: 10.1055/s-0031-1285829] [Medline: 21833896]
4. Jenabi E, Khazaei S, Bashirian S, Aghababaei S, Matinnia N. Reasons for elective cesarean section on maternal request: a systematic review. J Matern Fetal Neonatal Med 2020 Nov 08;33(22):3867-3872. [doi: 10.1080/14767058.2019.1587407] [Medline: 30810436]
5. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
6. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. nline health information seeking among US adults: measuring progress toward a Healthy People 2020 objective. Public Health Rep 2019;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]
7. Azer SA, Algrain HA, AlKhelaif RA, AlEshaiwi SM. Evaluation of the educational value of YouTube videos about physical examination of the cardiovascular and respiratory systems. J Med Internet Res 2013 Nov 13;15(11):e241 [FREE Full text] [doi: 10.2196/jmir.2728] [Medline: 24225171]
8. Koller U, Waldstein W, Schatz K, Windhager R. YouTube provides irrelevant information for the diagnosis and treatment of hip arthritis. International Orthopaedics (SICOT) 2016 Mar 31;40(10):1995-2002. [doi: 10.1007/s00264-016-3174-7]
9. Cassidy JT, Fitzgerald E, Cassidy ES, Cleary M, Byrne DP, Devitt BM, et al. YouTube provides poor information regarding anterior cruciate ligament injury and reconstruction. Knee Surg Sports Traumatol Arthrosc 2017 Mar 17;26(3):840-845. [doi: 10.1007/s00167-017-4514-x]
10. Erdem H, Sisik A. The reliability of bariatric surgery videos in YouTube platform. Obes Surg 2017 Sep 5;28(3):712-716. [doi: 10.1007/s11695-017-2911-3]
11. Google Adwords keyword planner. URL: https://adwords.google.com/KeywordPlanner [accessed 2021-04-02]
12. Williams D, Sullivan SJ, Schneiders AG, Ahmed OH, Lee H, Balasundaram AP, et al. Big hits on the small screen: an evaluation of concussion-related videos on YouTube. Br J Sports Med 2014 Jan;48(2):107-111. [doi: 10.1136/bjsports-2012-091853] [Medline: 23446643]
13. Hegarty E, Campbell C, Grammatopoulos E, DiBiase AT, Sherriff M, Cobourne MT. YouTube™ as an information resource for orthognathic surgery. J Orthod 2017 Jun;44(2):90-96. [doi: 10.1080/14653125.2017.1319010] [Medline: 28463076]
14. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
15. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb 01;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
16. Staunton PF, Baker JF, Green J, Devitt A. Online curves: a quality analysis of scoliosis videos on YouTube. Spine 2015;40(23):1857-1861. [doi: 10.1097/brs.0000000000001137]
17. Ferhatoglu MF, Kartal A, Ekici U, Gurkan A. Evaluation of the reliability, utility, and quality of the information in sleeve gastrectomy videos shared on open access video sharing platform YouTube. Obes Surg 2019 May;29(5):1477-1484. [doi: 10.1007/s11695-019-03738-2] [Medline: 30706318]

Abbreviations

VPI: video power index

Edited by R Kukafka; submitted 13.10.20; peer-reviewed by J Toelen, WC Lee; comments to author 01.03.21; revised version received 04.05.21; accepted 14.06.21; published 30.07.21.

Please cite as:
Lee KN, Joo YJ, Choi SY, Park ST, Lee KY, Kim Y, Son GH
Content Analysis and Quality Evaluation of Cesarean Delivery–Related Videos on YouTube: Cross-sectional Study
J Med Internet Res 2021;23(7):e24994
URL: https://www.jmir.org/2021/7/e24994
doi:10.2196/24994
PMID:34328422

XSL•FO
RenderX

©Kyong-No Lee, Yeon Ji Joo, So Yeon Choi, Sung Taek Park, Keun-Young Lee, Youngmi Kim, Ga-Hyun Son. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

US Privacy Laws Go Against Public Preferences and Impede Public Health and Research: Survey Study

Cason Schmit[1*], JD; Theodoros Giannouchos[1,2], PhD; Mahin Ramezani[1], BS; Qi Zheng[3], PhD; Michael A Morrisey[1], PhD; Hye-Chung Kum[1*], PhD, MSW

[1]Population Informatics Lab, Department of Health Policy and Management, Texas A&M University, College Station, TX, United States

[2]Pharmacotherapy Outcomes Research Center, College of Pharmacy, University of Utah, Salt Lake City, UT, United States

[3]Department of Epidemiology and Biostatistics, Texas A&M University, College Station, TX, United States

[*]these authors contributed equally

Corresponding Author:
Cason Schmit, JD
Population Informatics Lab
Department of Health Policy and Management
Texas A&M University
1266 TAMU
College Station, TX, 77843
United States
Phone: 1 (979)436 0277
Email: schmit@tamu.edu

Abstract

Background: Reaping the benefits from massive volumes of data collected in all sectors to improve population health, inform personalized medicine, and transform biomedical research requires the delicate balance between the benefits and risks of using individual-level data. There is a patchwork of US data protection laws that vary depending on the type of data, who is using it, and their intended purpose. Differences in these laws challenge big data projects using data from different sources. The decisions to permit or restrict data uses are determined by elected officials; therefore, constituent input is critical to finding the right balance between individual privacy and public benefits.

Objective: This study explores the US public's preferences for using identifiable data for different purposes without their consent.

Methods: We measured data use preferences of a nationally representative sample of 504 US adults by conducting a web-based survey in February 2020. The survey used a choice-based conjoint analysis. We selected choice-based conjoint attributes and levels based on 5 US data protection laws (Health Insurance Portability and Accountability Act, Family Educational Rights and Privacy Act, Privacy Act of 1974, Federal Trade Commission Act, and the Common Rule). There were 72 different combinations of attribute levels, representing different data use scenarios. Participants were given 12 pairs of data use scenarios and were asked to choose the scenario they were the most comfortable with. We then simulated the population preferences by using the hierarchical Bayes regression model using the ChoiceModelR package in R.

Results: Participants strongly preferred data reuse for public health and research than for profit-driven, marketing, or crime-detection activities. Participants also strongly preferred data use by universities or nonprofit organizations over data use by businesses and governments. Participants were fairly indifferent about the different types of data used (health, education, government, or economic data).

Conclusions: Our results show a notable incongruence between public preferences and current US data protection laws. Our findings appear to show that the US public favors data uses promoting social benefits over those promoting individual or organizational interests. This study provides strong support for continued efforts to provide safe access to useful data sets for research and public health. Policy makers should consider more robust public health and research data use exceptions to align laws with public preferences. In addition, policy makers who revise laws to enable data use for research and public health should consider more comprehensive protection mechanisms, including transparent use of data and accountability.

(J Med Internet Res 2021;23(7):e25266) doi:10.2196/25266

XSL•FO
RenderX

KEYWORDS

privacy; law; medical informatics; conjoint analysis; surveys and questionnaires; public health; information dissemination; health policy; public policy; big data

Introduction

Cleaning, integrating, and managing the uncertainty in chaotic real data is essential for reproducible science and to unleash the potential power of big data for biomedical research. This often requires access to very detailed data that inevitably raise privacy concerns. Despite the widespread use of personal information for big data purposes (eg, marketing, intelligence gathering, political campaigns), big data analytics are still challenged in health applications owing to concerns about privacy and complex and differing federal and state laws [1,2]. The patchwork of federal and state data protection laws poses a substantial challenge to leveraging data to promote health outcomes [1,2]. Data protection laws have 5 fundamental elements: (1) a definition of protected data, (2) definition of a regulated person or entity, (3) data use or disclosure restrictions, (4) data use or disclosure exceptions, and (5) penalties for violating the law. It is common for data protection laws to vary wildly in these 5 elements [2-6]. Consequently, it can be exceptionally difficult to understand which law (or laws) apply to a data project and whether a proposed data use is permitted. Often, the only commonality between different data protection laws is that they usually protect only identifiable data. Data that do not identify a person typically are not protected by US data protection laws. However, legal definitions for "identifiable" data or deidentification standards vary considerably [6]. This inconsistency encourages highly conservative measures to strip data of potential identifiers, which can severely limit data utility [6]. This reality poses a substantial barrier to cross-sectoral and cross-jurisdictional data uses relevant to health outcomes, including exploration of social determinants of health, retrospective database research studies, informatics research on decision support systems, digital ethology, and big data analytics in health (eg, precision public health) [2,7]. These barriers challenge efforts to rapidly leverage data in public health emergencies (eg, COVID-19).

An increasing number of published stories highlight the fact that different privacy protections apply in different contexts. For example, popular news stories have addressed how health information is treated differently when it is collected by health care providers as opposed to commercial companies such as Fitbit, Apple, or Ancestry.com [8,9]. Data use by health care providers is regulated by the Health Insurance Portability and Accountability Act of 1996 whereas data use by Fitbit or Apple is regulated by the Federal Trade Commission Act, which permits data use so long as they are neither unfair nor deceptive (ie, disclosed in a lengthy privacy policy) [10].

Recent high profile breaches (eg, Equifax) and scandals (eg, Facebook and the 2016 US election) have raised awareness of these different privacy standards [10,11]. Moreover, new data protection regulations in some jurisdictions have provoked debate and congressional inquiry into new federal privacy legislation [12-15]. Any new federal privacy law will necessarily address the 5 fundamental elements of data protection laws and will inevitably address the trade-off between privacy and utility [16,17]. Privacy risk in database studies is best minimized through a holistic approach that involves security technology (eg, encryption), data manipulation (eg, differential privacy), as well as good data governance models (eg, transparency) and legal protections. Legal protections can shield against a variety of harms. Alternatively, permitting certain data use can promote social benefits, including advancing knowledge and science, promoting public health, facilitating law enforcement, and enabling economic activity [18]. In the United States, the decision to permit or restrict certain data uses is determined by elected officials. These are policy choices with significant consequences for both individual interests (eg, privacy) and public benefit. Consequently, constituent input is critical to finding the right balance between individual privacy and public benefits.

The purpose of this paper is to report on the results of a nationally representative survey examining US residents' preferences for which of their identifiable personal data should be available for use, by whom, and for what purposes. Prior research focusing on Americans' attitudes on data use and privacy shows strong support for socially beneficial uses such as research [19-23]. However, few US privacy laws provide specific exceptions for data uses for research or public health [1,2]. Thus, information on how the US population views certain data uses in relation to other data uses is valuable, especially if one data use is currently restricted under US laws and the other is permitted. Such data would be extremely useful to US policy makers as they deliberate new data protection frameworks.

Methods

Study Design and Recruitment

In February 2020, we conducted a web-based survey to explore the comfort levels and the preferences of the US population when individually identifiable data is reused for different purposes without their consent. Potential participants were recruited via a third-party research company (Dynata) that specializes in deploying surveys by using nationally representative sampling. We sought to balance the sample on 6 targets based on population characteristics used by the census (gender, race/ethnicity, age, education, household income, and region) where possible. Our goal was to recruit 500 adult (≥18 years) US residents fluent in English to enable reasonable sample balancing [24]. To provide a demographic context for participants' baseline privacy concerns, we included the validated Concern for Information Privacy instrument [25,26]. The Concern for Information Privacy instrument has 15 seven-point Likert scale questions (1=strongly disagree to 7=strongly agree) and provides a composite score with 4 subscales for privacy [25]. We estimated participants' preferences on the potential reuse of their data with a choice-based conjoint analysis [27-30]. Choice-based conjoint analysis is built on the premise that an individual places different

XSL•FO
RenderX

values on an option according to its characteristics and makes trade-off choices among alternatives based on the combination of characteristics. Conjoint analysis is a decomposition method because the implicit value for a characteristic is derived from some overall score for a profile consisting of 2 or more (conjoint) characteristics [27-29,31]. Choice-based conjoint analysis is commonly used in health care and economics research to understand clinical decision making, to assess patients' preferences, evaluation, and willingness to accept new treatments and health states, and to promote shared decision making among patients and stakeholders by quantifying the relative importance that individuals place on different attributes and levels within those attributes [27-30,32-34].

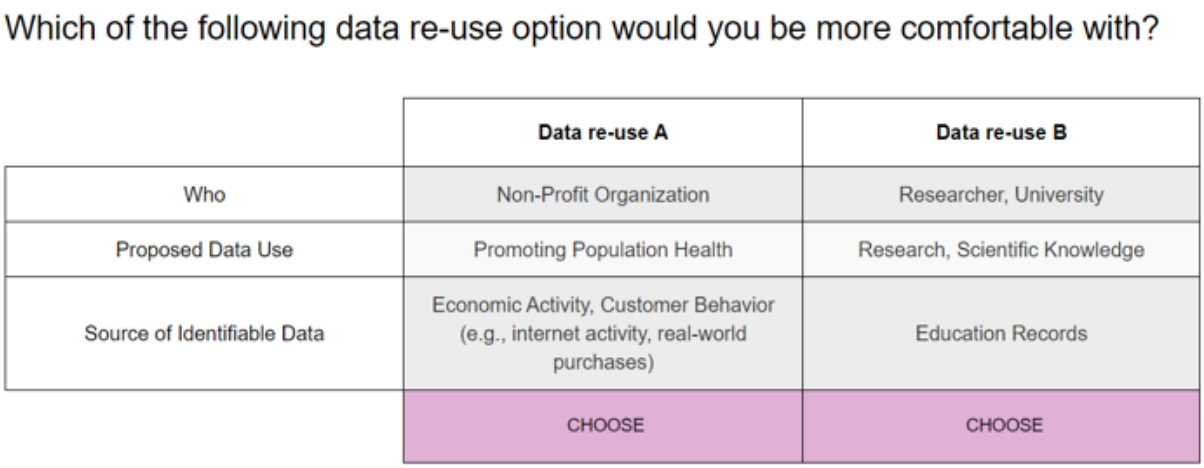
We selected attributes based on 4 of the 5 elements of the data protection laws (excluding violation penalties) (Table 1). The "source of identifiable data" is related to the definition of protected data, "who" is related to the regulated entity, and the "proposed data use" is related to 2 different elements: legal restrictions and exceptions for data use or disclosure [2]. We selected attribute levels to correspond to various legal provisions permitting or restricting data reuse and to identify the nuances within the categories (eg, business vs nonprofit organization), which resulted in 80 different scenarios comprising different attribute levels (4×4×5) [10,35-38]. Of those, we excluded 8 as implausible or likely to confuse survey respondents (eg, government or nonprofit conducting profit-driven activity), leaving 72 different scenarios.

Table 1. Attributes and levels for data reuse scenarios.

Who	Purpose	Source of identifiable data
Researcher, University	Research, scientific knowledge dissemination	Education records
Nonprofit Organization	Promoting population health	Health records
Government	Identify criminal activity	Government program or activity
Business	Marketing, recruitment Profit-driven activity	Economic activity, customer behavior

Since it is not feasible and manageable to present all the possible combinations of each scenario to the participants, a fractional factorial design was used to randomly generate subsets of all the combinations, which were sufficient to obtain robust and meaningful differences in preferences through a standard web-based platform called "conjoint.ly", similar to that reported in previous work [33,34,39,40]. This resulted in 72 choice sets, with each set consisting of 12 pairs of data use scenarios that would allow for simulating participant preferences in the full space of data use scenario permutations. Each participant was randomly assigned to respond to one of the choice sets, and we asked each participant to select the data use scenario that they were the most comfortable with for each of their assigned 12 scenario pairs (Figure 1).

Figure 1. Sample pair scenario question.



Statistical Analysis

To estimate the parameters, we used a hierarchical Bayes regression model, and in estimating the parameters at the individual level, we generated 10,000 posterior draws by using the Markov chain Monte Carlo simulation [29,41,42]. This approach allowed the estimation of attributes and levels with the small amount of data collected from each respondent, while simultaneously accounting for the heterogeneity of preferences across and within individuals, the nested structure of the choices, and thus, the nonrandom preference variation of the respondents [29,34,41-43]. The value of 3 attributes was scaled to sum up to 100%, while the values of the levels within each attribute (part worth utilities) sum up to zero, with negative values

XSL•FO
RenderX

indicating decreased and positive values indicating increased preferences. Finally, we used the sum of the estimated relative values (utilities) of different levels to identify and rank the alternative scenarios from the most to the least preferable. All measures were estimated at the individual level, which were then averaged and reported as the population mean with standard deviation in the results. Data analyses were conducted using the ChoiceModelR package in R [44]. This study was approved by our university institutional review board.

Results

The survey was distributed to 687 individuals. Of them, 22 individuals declined to participate (3.2%), 157 did not fully complete the survey (22.8%), and 4 participant responses (0.6%) were marked as low quality based on detected participant behavior (eg, rapidly clicking through without mouse movement). This resulted in 504 respondents who fully completed the web-based survey (response rate 74.4%), which was our final analytic sample. Generally, we were able to meet our census sampling targets for gender, race/ethnicity, age, education, income, and census region (Table 2). In addition, although we did not try to balance for it, our sample's health insurance coverage is similar to data published by the US Census Bureau [45]. Around half of the respondents had used a health care provider in the past year, around one-third had at least one chronic condition, and around 19.8% (100/504) of the respondents visited an emergency department in the past year. The overall privacy score was 5.8 (SD 1.1), which is consistent with the Concern for Information Privacy validation samples (ie, scores ranging from 5 to 6) [25].

XSL•FO
RenderX

Table 2. Sociodemographic data, clinical characteristics, and privacy attitude scores of the participants (N=504).

Participant characteristics	Values	Target sample percentage[a,b]
Age categories (years), n (%)		
18-24	41 (8.1)	13.1
25-34	75 (14.9)	17.5
35-44	100 (19.8)	17.5
45-54	101 (20.0)	19.2
55-64	68 (13.5)	15.6
65 or older	89 (17.7)	17.2
Gender, n (%)		
Male	224 (44.4)	48.5
Female	278 (55.2)	50.5
Other/prefer not to answer	2 (0.4)	—[c]
Race categories, n (%)		
White	315 (62.5)	63.7
African American	77 (15.3)	12.2
Hispanic	51 (10.1)	16.4
Asian	46 (9.1)	4.7
Other	15 (3.0)	3.0
Income categories, n (%)		
$20,000 or less	103 (20.4)	19.9
$20,000 to $49,999	149 (29.6)	30.6
$50,000 to $99,999	137 (27.2)	29.1
$100,000 to $149,999	67 (13.3)	12.0
$150,000 or more	48 (9.5)	8.3
Educational level, n (%)		
High school or less	172 (34.1)	32.0
Some college completed	99 (19.6)	19.0
College degree	191 (37.9)	31.0
Master's	37 (7.3)	—
PhD/doctoral	5 (1.0)	—
Region, n (%)		
Midwest	95 (18.8)	22.0
Northeast	126 (25.0)	18.2
South	174 (34.5)	36.2
West	109 (21.6)	23.6
Health insurance coverage, n (%)[b]		
Private	169 (33.5)	64.7
Medicare	112 (22.2)	17.7
Medicaid	83 (16.5)	17.9
Uninsured	52 (10.3)	8.5
VA/TRICARE	10 (2.0)	3.6
Multiple	78 (15.5)	14.5

XSL•FO
RenderX

Participant characteristics	Values	Target sample percentage[a,b]
Any chronic condition, n (%)		
No	319 (63.3)	—
Yes	181 (35.9)	—
Use of health care provider in the past year, n (%)		
No	93 (18.5)	—
Yes	256 (50.8)	—
At least one emergency department visit in the past year, n (%)		
No	404 (80.2)	—
Yes	100 (19.8)	—
Respondent is a primary care giver for someone else, n (%)		
No	423 (83.9)	—
Yes	77 (15.3)	—
Concern for information privacy scores, mean (SD)	5.8 (1.1)	—

[a]Survey sampling targets based on census data.

[b]Insurance data were not used as the sampling target. These data show 2018 insurance statistics from the US census for survey sampling comparisons [45]. Our survey solicited mutually exclusive responses in contrast to the US census data, which do not exclude persons with multiple insurance types from these groups.

[c]Not available.

Figure 2 presents the relative importance for the different levels within each attribute. Positive values indicate preference with higher values and reveal greater importance, while negative values indicate nonpreferred levels associated with potential data reuse. Participants were most comfortable with the reuse of identifiable data if the proposed data use was intended to promote population health (10.1%, SD 11.6) or promote science or research (8.2%, SD 6.5), if the data were used by university-affiliated researchers (6.4%, SD 10.7) or nonprofit organizations (2.5%, SD 16.1), and if the source of the data included educational (2.2%, SD 11.3) or health care records (1.4%, SD 10.4). In contrast, participants were least comfortable with data reuse by businesses (–4.5%, SD 13.7) or the government (–4.3%, SD 16.8) mainly for profit-driven (–11.7%, SD 12.3) or marketing (–4.2%, SD 11) activities based on governmental (–1.7%, SD 10.1) or economic activity data (–1.8%, SD 11.4). Overall, we observed higher differences in the values between the levels of the proposed data use attribute compared to other 2 attributes, particularly with the attribute related to the source of the identifiable data.

XSL•FO
RenderX

Figure 2. Relative importance by level within each attribute in percentage (SD).

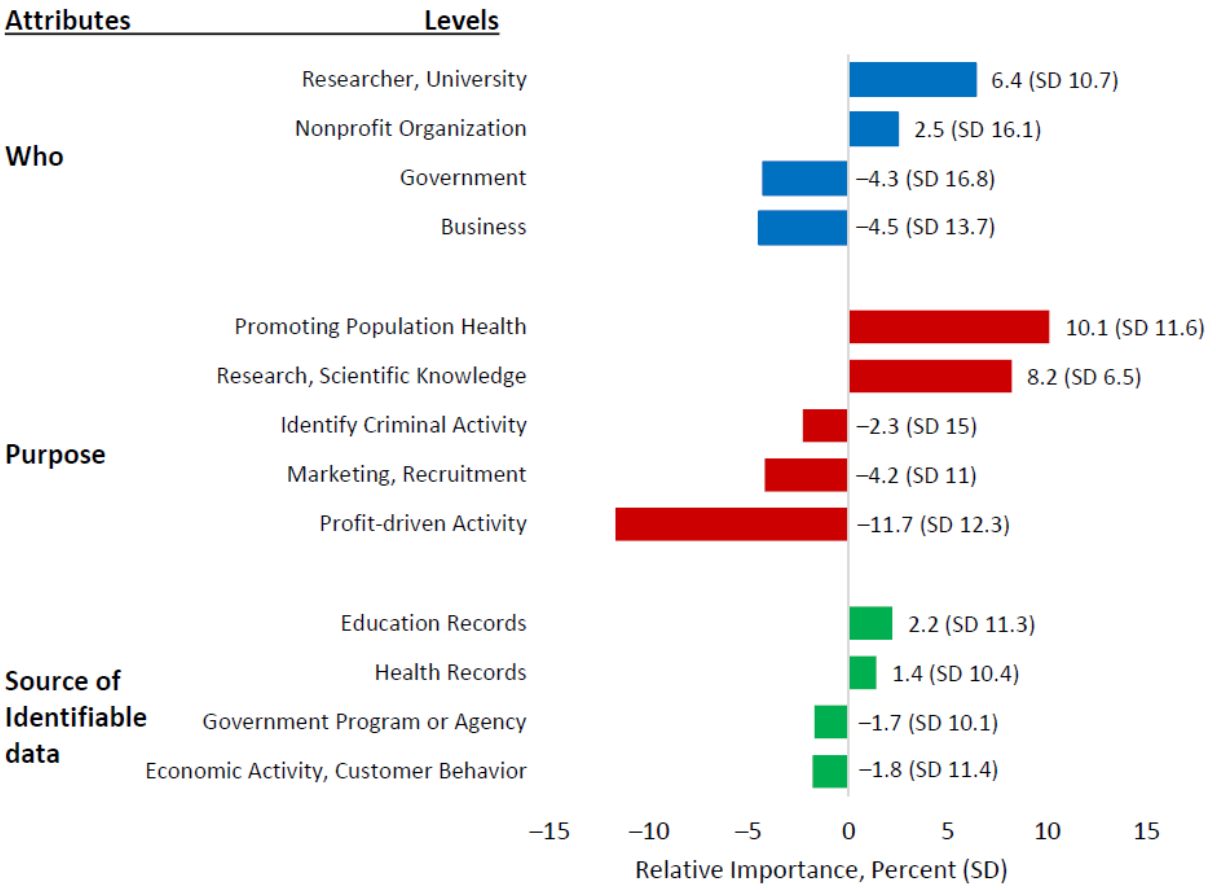
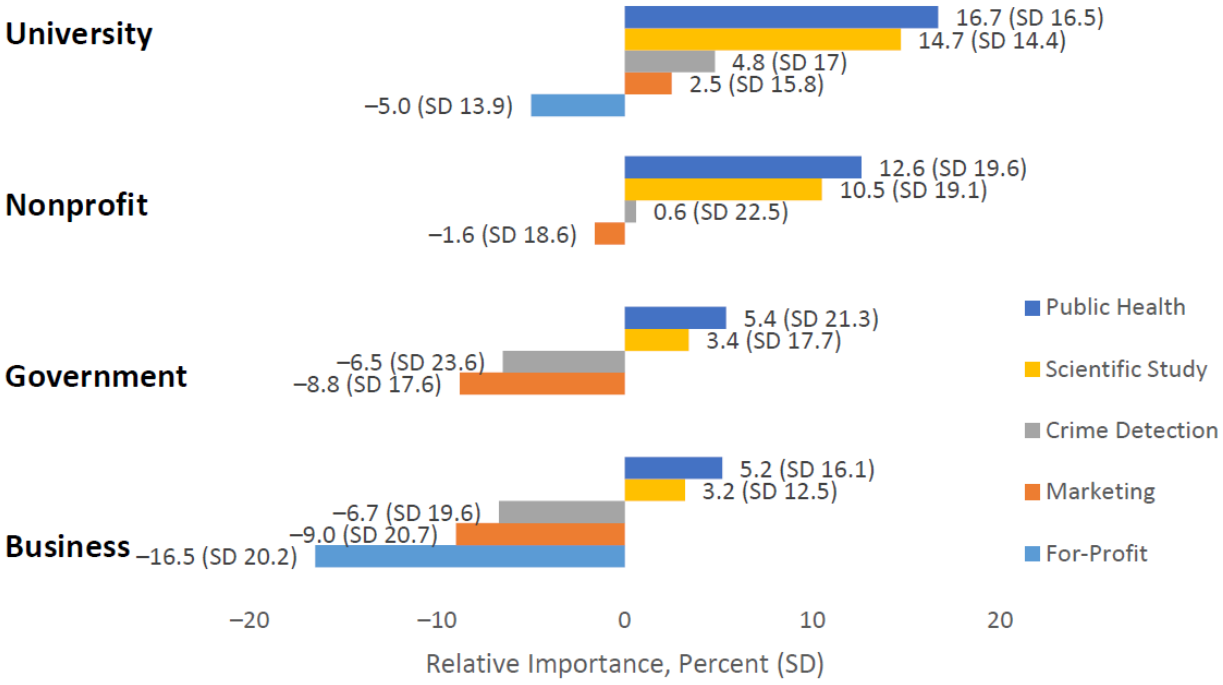


Figure 3 presents the relative importance of the different data use purposes by users. All data use activities were more preferred when conducted by universities or nonprofit organizations than when conducted by government or business. Public health and research activities received positive relative importance values regardless of who conducts the activity. Conversely, for-profit and marketing activities received near uniform negative relative importance values. Universities conducting marketing activities was the lone exception, which had a small but positive relative importance value (2.6%). The relative importance of identifying criminal activity was either positive (university and nonprofit) or negative (government and business) depending on the user. We also estimated the overall values that participants placed on each scenario. Figure 4 presents only the 10 highest and lowest ranked scenarios (Figure 4). The 4 lowest ranked scenarios all involved businesses using data for profit-driven purposes. The remainder of the lowest ranked scenarios involved business or government organizations engaging in marketing or identifying criminal activity. Eight of the 10 highest ranked scenarios involved universities/researchers engaging in scientific research or public health activities. Nonprofit organizations conducting population health programs represented the seventh and eighth highest ranked scenarios. We checked model validity by comparing the actual choices made by each participant with the estimated choices made for at least 90% of the last 50 iterations of the Markov chain Monte Carlo simulation. Precision (% of correct estimates) was good at 92% for the simulated model.

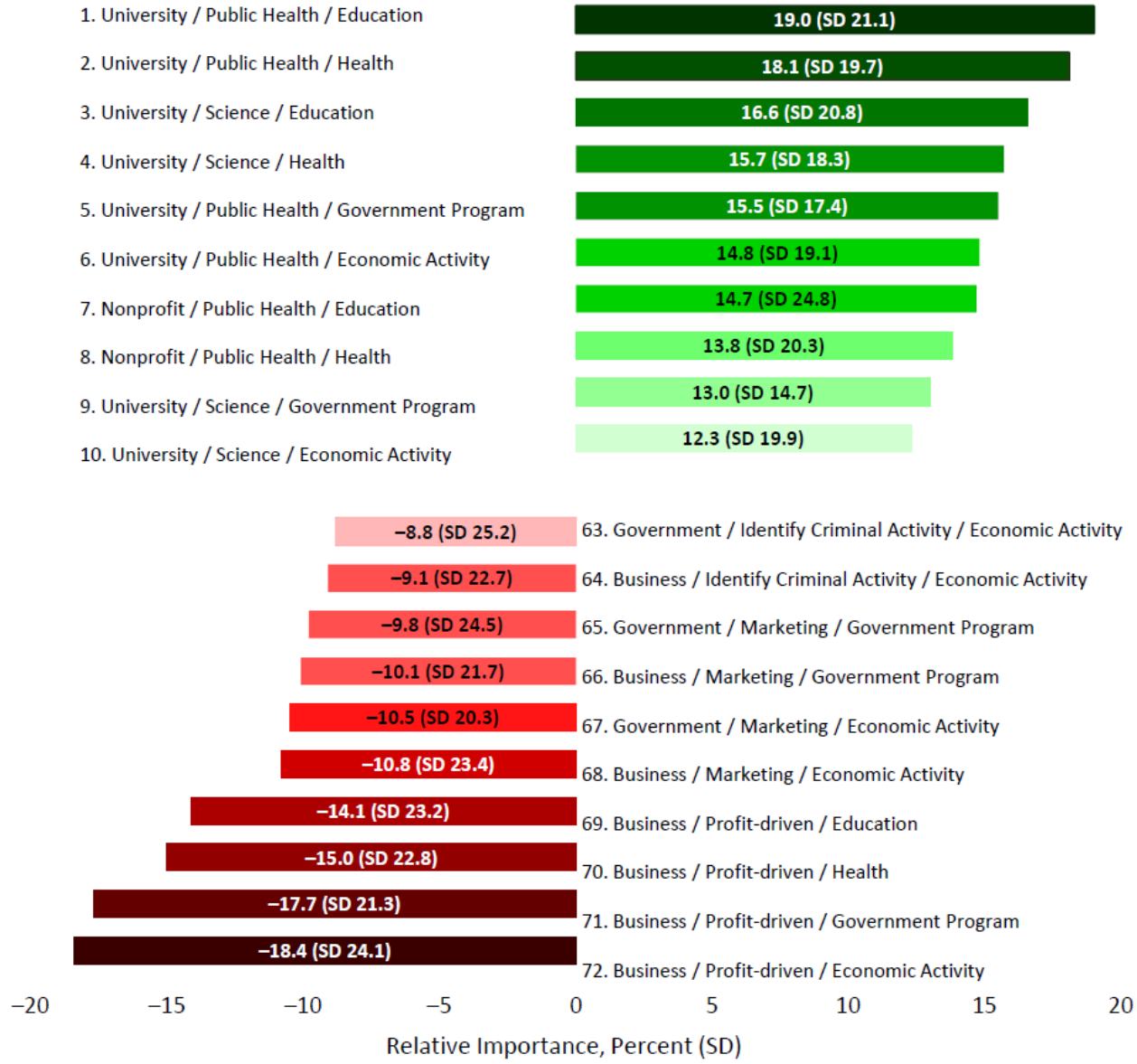
XSL•FO
RenderX

Figure 3. Public preferences for use of data by users and purpose in percentage (SD). Our survey did not pair "for-profit" purposes with government or nonprofit users because these pairings were implausible and likely to confuse survey respondents.



XSL•FO
RenderX

Figure 4. Top 10 and bottom 10 ranked data use scenarios derived from the sum of scenario attributes' relative values (who/use purpose/data source) in percentage (SD).



Discussion

Principal Findings

In contrast to federal and state laws, US residents make little distinctions across types of data. However, they express much more favorable preferences for uses by academic researchers and nonprofit organizations than by the government or the business community. Moreover, all types of users consistently preferred uses that focus on public health and scientific research rather than on crime detection, marketing, or for-profit activities. Our data demonstrate interesting inconsistencies between public preferences and US privacy laws. These inconsistencies are best exemplified by our participants' most preferred data reuse (researchers using education data to promote population health) and least preferred data reuse (businesses using consumer data for profit). Ironically, our data indicate that the US public's most preferred data reuse scenario is currently prohibited under the federal Family Educational Rights and Privacy Act of 1974 while the US public's least preferred data reuse is completely legal and ubiquitous under the permissive Federal Trade Commission Act [38,46]. A recent review of 22 federal data protection frameworks funded by the Network for Public Health Law indicates that few data protection laws have a general research exception and fewer have a specific exception for public health uses. However, these data uses were by far the most preferred options of those we presented to our participant and were consistently preferred, regardless of who was the data user. Yet, public health uses are treated differently under different US laws [6]. For example, the law protecting substance use treatment records has hamstrung the use of data in the present opioid epidemic response, while the laws covering cell phone location data have permitted public health officials to track entire populations in the current COVID-19 pandemic [47,48].

XSL•FO
RenderX

Our participants also strongly favored data uses by universities and nonprofit organizations. Both universities and nonprofit organizations received higher preference ratings for all data use activities when compared to those received by the government or businesses. In some cases, activities that participants viewed as heavily undesirable when conducted by the government or business (crime detection, marketing) were rated favorably when conducted by a nonprofit organization or university. In contrast, the least preferred scenarios involved data reused for profit-driven or marketing activities by businesses or government. Mistrust in government has been documented in other research on attitudes of research and is perhaps unsurprising in the present partisan political environment [21]. Negative preference ratings for businesses, profit-driven activities, and marketing are likely due to frequent stories of controversial data use, mismanagement, or breaches that are all too common in the news [49]. This finding is consistent with other research documenting strong public attitudes in favor of altruistic goals and skepticism of data uses that advance specific individual or private (ie, for-profit) interests [19,21].

We did find some preference differences for certain data types, but these differences were modest. Our data show that the public prefers the use of health or educational data (both heavily regulated under US laws) as compared to government data or economic data. Still, our data do not show any strong preferences. The public seems to view data as data. We noted that 4 of the 5 data use purposes we included in our study fall neatly into 2 broad categories: altruistic purposes and self-serving purposes. Public health and scientific purposes both ultimately contribute to the greater good, and our data suggest that these purposes are strongly preferred by the US public, regardless of who is doing the activity. In contrast, our respondents generally found those activities that are primarily self-serving (ie, profit-driven or marketing/recruitment activities) undesirable, regardless of who was doing the activity. The lone exception was marketing by universities, which received a modest positive relative importance score. Consequently, it could be that our participants based some of their preference decisions on whether they saw the data use as contributing to an altruistic or common good objective as opposed to primarily benefiting the data user's self-interests.

Identifying criminal activity was the one data use that does not neatly fit in the broad categories of altruistic or self-serving purposes. While law enforcement clearly has some social benefits (as do all the activities used in our study), identifying criminal activity implies punishment for some individuals. Consequently, it is not entirely altruistic and not entirely self-serving. Interestingly, participant preferences for identifying criminal activity seemed to vary depending on the data user. Universities and nonprofit organizations both received positive relative importance scores whereas governments and businesses received negative scores. Just as with other data uses, it could be that participants positively associate universities and nonprofit organizations with motivations more in line with social benefits rather than individual benefits.

Collectively, our results do not support the current patchwork of US data protection laws. Many US data protection laws focus primarily on the type of data (ie, health, education, governmental program data), but our respondents were fairly indifferent toward these distinctions. Instead, our findings suggest that the US public is much more interested in who is using the data and for what purposes the data are being used. In particular, our results suggest that the US public has a strong preference for data uses that promote the common good as opposed to individual or self-serving interests.

In fact, findings suggest that US preferences more closely align with a comprehensive data protection framework such as the General Data Protection Regulation enacted by the European Union where rules vary based on data use but are broadly applicable to all identifiable data [50]. For example, the General Data Protection Regulation has broad applications and express provisions permitting scientific research and activities in the public interest (eg, public health) [51,52]. Policy makers who revise laws to increase access to data for research and public health can support data protection through new security standards. A 2009 report by the Institute of Medicine, "Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health through Research" [53] argued for a different data protection approach to "enhance privacy protections through improved data security, increased transparency of activities and policies, and greater accountability." These good governance practices, as opposed to strict prohibitions on uses and disclosures (ie, for research or public health), provide a way to protect individuals while permitting big data applications (eg, linking data from different sources) with social benefits. These results provide strong public support for continued efforts to make data available for research and public health.

Limitations

There are 2 important limitations. We did not capture the universe of data use possibilities; therefore, the measured participants' preferences are relative to the 72 provided scenarios. Additionally, this design measured participants' preferences rather than acceptability, meaning that a participants' least preferred scenario could still be acceptable to them or the most preferred scenario might be unacceptable.

Conclusion

Importantly, these results support a close re-examination of the absence of public health and research data use exceptions in US laws. It is clear that the US public strongly prefers using data to promote population health (as compared to other legal data uses); yet, few laws allow this kind of exception. The Family Educational Rights and Privacy Act provides an excellent example, given that it does not have a public health exception (or a research exception that permits exploring health implications) despite being one of the most potent known social determinants of health. Moreover, the absence of these data use exceptions within the current patchwork of inconsistent US data protection laws persistently frustrates secondary database researchers and public health professionals, thereby delaying, impeding, or increasing the cost of data-intensive scientific discovery and public health practice [1,2,4,6]. These findings clearly show that there is poor alignment between the present US legal data protection framework and the preferences of the US population.

XSL•FO
RenderX

Acknowledgments

This study was funded in part by the Presidential Impact Fellow Award at Texas A&M University and the Population Informatics Lab in the School of Public Health at Texas A&M University.

Authors' Contributions

CS and HCK were responsible for conceptualization, study design and implementation, data analysis, interpretation, writing and revisions, and supervision. TG and MR were responsible for study implementation, data analysis, interpretation, writing, and revisions. QZ and MM were responsible for data analysis, interpretation, writing, and revisions.

Conflicts of Interest

None declared.

References

1. Hulkower R, Penn M, Schmit C. Privacy and confidentiality of public health information. In: Magnuson JA, Dixon BE, editors. Public Health Informatics and Information Systems. Cham, Switzerland: Springer Nature; 2020:147-166.
2. Schmit C, Kelly K, Bernstein J. Cross Sector Data Sharing: Necessity, Challenge, and Hope. J Law Med Ethics 2019 Jun;47(2_suppl):83-86. [doi: 10.1177/1073110519857325] [Medline: 31298141]
3. Begley EB, Ware JM, Hexem SA, Rapposelli K, Thompson K, Penn MS, et al. Personally Identifiable Information in State Laws: Use, Release, and Collaboration at Health Departments. Am J Public Health 2017 Aug;107(8):1272-1276. [doi: 10.2105/ajph.2017.303862]
4. O'Connor J, Matthews G. Informational Privacy, Public Health, and State Laws. Am J Public Health 2011 Oct;101(10):1845-1850. [doi: 10.2105/ajph.2011.300206]
5. Savage L. To bring health information privacy into the 21st century, look beyond HIPAA. Health Affairs Blog. 2018 Jul 05. URL: https://www.healthaffairs.org/do/10.1377/hblog20180702.168974/full/ [accessed 2020-07-15]
6. Wartenberg D, Thompson WD. Privacy Versus Public Health: The Impact of Current Confidentiality Rules. Am J Public Health 2010 Mar;100(3):407-412. [doi: 10.2105/ajph.2009.166249]
7. Kum HC, Ahalt S. Privacy-by-Design: Understanding Data Access Models for Secondary Data. AMIA Jt Summits Transl Sci Proc 2013;2013:126-130 [FREE Full text] [Medline: 24303251]
8. Elfin D. DNA testing? you might want to wait for more legal protection. Bloomberg Law. 2019 Jan 07. URL: https://news.bloomberglaw.com/pharma-and-life-sciences/dna-testing-you-might-want-to-wait-for-more-legal-protection [accessed 2020-05-25]
9. North T. Google, Fitbit, and the sale of our private health data. Health Law & Policy Brief. 2019 Nov 18. URL: http://www.healthlawpolicy.org/google-fitbit-and-the-sale-of-our-private-health-data/ [accessed 2020-05-25]
10. Federal Trade Commission Act. Federal Trade Commission. 2012. URL: https://www.ftc.gov/enforcement/statutes/federal-trade-commission-act [accessed 2021-06-14]
11. Granville K. Facebook and Cambridge Analytica: What you need to know as the fallout widens. The New York Times. 2018 Mar 19. URL: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html [accessed 2020-05-05]
12. California Consumer Privacy Act. State of California Department of Justice. 2018. URL: https://oag.ca.gov/privacy/ccpa [accessed 2021-06-14]
13. European Commission. General Data Protection Regulation. Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0249&qid=1623685068702 [accessed 2021-06-14]
14. Cameron D. FTC members unanimously press Congress for tough national privacy protections. Gizmodo. 2019 May 08. URL: https://gizmodo.com/ftc-members-unanimously-press-congress-for-tough-nation-1834624258 [accessed 2020-05-21]
15. Mulligan S, Freeman W, Linebaugh C. Data protection law: An overview. Congressional Research Service. URL: https://crsreports.congress.gov/product/pdf/R/R45631 [accessed 2020-05-06]
16. Dinur I, Nissim K. Revealing Information while Preserving Privacy. PODS '03: Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems 2003 Jun 09:202-210 [FREE Full text] [doi: 10.1145/773153.773173]
17. Ragan E, Kum H, Ilangovan G, Wang H. Balancing privacy and information disclosure in interactive record linkage with visual masking. 2018 Apr 20 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; April; New York p. 1-12. [doi: 10.1145/3173574.3173900]
18. Kum H, Krishnamurthy A, Machanavajjhala A, Ahalt SC. Social Genome: Putting Big Data to Work for Population Informatics. Computer 2014 Jan;47(1):56-63. [doi: 10.1109/mc.2013.405]

XSL•FO
RenderX

19. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
20. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
21. Haga S, O'Daniel J. Public perspectives regarding data-sharing practices in genomics research. Public Health Genomics 2011;14(6):319-324 [FREE Full text] [doi: 10.1159/000324705] [Medline: 21430368]
22. O'Brien EC, Rodriguez AM, Kum H, Schanberg LE, Fitz-Randolph M, O'Brien SM, et al. Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire. Int J Med Inform 2019 Jul;127:9-17 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.04.003] [Medline: 31128837]
23. Willison D, Keshavjee K, Nair K, Goldsmith C, Holbrook AM, Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness investigators. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. BMJ 2003 Feb 15;326(7385):373 [FREE Full text] [doi: 10.1136/bmj.326.7385.373] [Medline: 12586673]
24. Siddiqui K. Heuristics for sample size determination in multivariate statistical techniques. World Applied Sciences Journal 2013;27(2):285-287. [doi: 10.1002/0471667196.ess2329]
25. Smith HJ, Milberg SJ, Burke SJ. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly 1996 Jun;20(2):167. [doi: 10.2307/249477]
26. Stewart KA, Segars AH. An Empirical Examination of the Concern for Information Privacy Instrument. Information Systems Research 2002 Mar;13(1):36-49. [doi: 10.1287/isre.13.1.36.97]
27. Ben-Akiva M, de Palma A, McFadden D, Abou-Zeid M, Chiappori P, de Lapparent M, et al. Process and context in choice models. Mark Lett 2012 May 17;23(2):439-456. [doi: 10.1007/s11002-012-9180-7]
28. Ben-Akiva M, McFadden D, Abe M. Modeling methods for discrete choice analysis. Marketing Letters. 1997. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/47225/11002_2004_Article_138116.pdf?sequence=1 [accessed 2021-06-14]
29. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health 2011 Jun;14(4):403-413 [FREE Full text] [doi: 10.1016/j.jval.2010.11.013] [Medline: 21669364]
30. Ryan M, Farrar SJB. Using conjoint analysis to elicit preferences for health care. BMJ 2000 Jun 03;320(7248):1530-1533 [FREE Full text] [doi: 10.1136/bmj.320.7248.1530] [Medline: 10834905]
31. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26(8):661-677. [doi: 10.2165/00019053-200826080-00004] [Medline: 18620460]
32. Marshall D, Bridges JF, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008. Patient 2010 Dec 01;3(4):249-256. [doi: 10.2165/11539650-000000000-00000] [Medline: 22273432]
33. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics 2019 Feb;37(2):201-226 [FREE Full text] [doi: 10.1007/s40273-018-0734-2] [Medline: 30392040]
34. Zimmermann TM, Clouth J, Elosge M, Heurich M, Schneider E, Wilhelm S, et al. Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. J Affect Disord 2013 Jun;148(2-3):210-219. [doi: 10.1016/j.jad.2012.11.062] [Medline: 23290792]
35. 45 CFR Parts 160 and 164. HIPAA Privacy Rule. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/introduction/index.html [accessed 2021-06-14]
36. Lamdan S. Privacy act of 1974. The SAGE Encyclopedia of Surveillance, Security, and Privacy. 2018. URL: http://sk.sagepub.com/reference/the-sage-encyclopedia-of-surveillance-security-privacy/i9567.xml [accessed 2021-06-14]
37. Porter JP. The Federal Policy for the Protection of Human Subjects. IRB: Ethics and Human Research 1991 Sep;13(5):8. [doi: 10.2307/3564127]
38. Family educational rights and privacy act (FERPA). US Department of Education. 1988. URL: https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html [accessed 2021-06-14]
39. Frost N, Freeman J, Brixner D, Mort J, Clem J, Ngorsuraches S. Patients' preferences and willingness-to-pay for disease-modifying therapies. Mult Scler Relat Disord 2019 Oct;35:55-60. [doi: 10.1016/j.msard.2019.07.005] [Medline: 31323483]
40. Jaynes J, Wong W, Xu H. Using blocked fractional factorial designs to construct discrete choice experiments for healthcare studies. Stat Med 2016 Jul 10;35(15):2543-2560 [FREE Full text] [doi: 10.1002/sim.6882] [Medline: 26823156]
41. Allenby G, Rossi P. Hierarchical bayes models. In: Grover R, Vriens M, editors. The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Thousand Oaks, CA: Sage Publications, Inc; 2006.
42. Mohammadi T, Zhang W, Sou J, Langlois S, Munro S, Anis AH. A Hierarchical Bayes Approach to Modeling Heterogeneity in Discrete Choice Experiments: An Application to Public Preferences for Prenatal Screening. Patient 2020 Apr;13(2):211-223. [doi: 10.1007/s40271-019-00402-w] [Medline: 31814082]

XSL•FO
RenderX

43. Regier DA, Ryan M, Phimister E, Marra CA. Bayesian and classical estimation of mixed logit: An application to genetic testing. J Health Econ 2009 May;28(3):598-610. [doi: 10.1016/j.jhealeco.2008.11.003] [Medline: 19345433]
44. Sermas R, Colias J. ChoiceModelR. 2015. URL: https://cran.r-project.org/web/packages/ChoiceModelR/ChoiceModelR.pdf [accessed 2020-05-06]
45. Berchick E, Hood E, Barnett J. Health insurance coverage in the United States: 2018. US Census Bureau. 2019 Nov. URL: https://www.census.gov/content/dam/Census/library/publications/2019/demo/p60-267.pdf [accessed 2021-06-14]
46. Solove DJ, Hartzog W. The FTC and the New Common Law of Privacy. Columbia Law Review 2014;114:583. [doi: 10.2139/ssrn.2312913]
47. Frakt AB, Bagley N. Protection or Harm? Suppressing Substance-Use Data. N Engl J Med 2015 May 14;372(20):1879-1881. [doi: 10.1056/nejmp1501362]
48. O'Sullivan D. How the cell phones of spring breakers who flouted coronavirus warnings were tracked? CNN. 2020 Apr 04. URL: https://www.cnn.com/2020/04/04/tech/location-tracking-florida-coronavirus/index.html [accessed 2020-05-25]
49. Equifax data breach settlement. Federal Trade Commission. 2020. URL: https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement [accessed 2020-05-06]
50. Bari L, O'Neill D. Rethinking patient data privacy in the era of digital health. Health Affairs Blog. 2019. URL: https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/ [accessed 2020-07-16]
51. European Parliament. GDPR article 6 (1)(e). Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2021-06-14]
52. European Parliament. GDPR articles 5 (1)(b), 89(1). Official Journal of the European Union. 2018 May 25. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2021-06-14]
53. Nass S, Levit L, Gostin L, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington DC: National Academies Press; 2009.

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by J Jose, P Swire; comments to author 22.11.20; revised version received 08.12.20; accepted 30.04.21; published 05.07.21.

Please cite as:
Schmit C, Giannouchos T, Ramezani M, Zheng Q, Morrisey MA, Kum HC
US Privacy Laws Go Against Public Preferences and Impede Public Health and Research: Survey Study
J Med Internet Res 2021;23(7):e25266
URL: https://www.jmir.org/2021/7/e25266
doi:10.2196/25266
PMID:36260399

©Cason Schmit, Theodoros Giannouchos, Mahin Ramezani, Qi Zheng, Michael A Morrisey, Hye-Chung Kum. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Hospital Productivity After Data Breaches: Difference-in-Differences Analysis

Jinhyung Lee[1,2], PhD; Sung J Choi[3], PhD

[1]Department of Economics, Sungkyunkwan University, Seoul, Republic of Korea

[2]Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, United States

[3]School of Global Health Management and Informatics, College of Community Innovation and Education, University of Central Florida, Orlando, FL, United States

Corresponding Author:
Sung J Choi, PhD
School of Global Health Management and Informatics
College of Community Innovation and Education
University of Central Florida
528 West Livingston St
Orlando, FL, 32801
United States
Phone: 1 407 823 2369
Email: sung.choi@ucf.edu

Abstract

Background: Data breaches are an inevitable risk to hospitals operating with information technology. The financial costs associated with data breaches are also growing. The costs associated with a data breach may divert resources away from patient care, thus negatively affecting hospital productivity.

Objective: After a data breach, the resulting regulatory enforcement and remediation are a shock to a hospital's patient care delivery. Exploiting this shock, this study aimed to investigate the association between hospital data breaches and productivity by using a generalized difference-in-differences model with multiple prebreach and postbreach periods.

Methods: The study analyzed the hospital financial data of the California Office of Statewide Health Planning and Development from 2012 to 2016. The study sample was an unbalanced panel of hospitals with 2610 unique hospital-year observations, including general acute care hospitals. California hospital data were merged with breach data published by the US Department of Health and Human Services. The dependent variable was hospital productivity measured as value added. The difference-in-differences model was estimated using fixed effects regression.

Results: Hospital productivity did not significantly differ from the baseline for 3 years after a breach. Data breaches were not significantly associated with a reduction in hospital productivity. Before a breach, the productivity of hospitals that experienced a data breach maintained a parallel trend with control hospitals.

Conclusions: Hospital productivity was resilient against the shocks from a data breach. Nonetheless, data breaches continue to threaten hospitals; therefore, health care workers should be trained in cybersecurity to mitigate disruptions.

(J Med Internet Res 2021;23(7):e26157) doi:10.2196/26157

KEYWORDS

cybersecurity; data breach; health information technology; health information; hospital data breach; hospital productivity; information technology; privacy

Introduction

Data breaches are an inevitable risk to hospitals operating with information technology (IT). The US Department of Health and Human Services (HHS) defines a data breach as the impermissible use or disclosure of protected health information [1] and can be categorized as follows: theft, loss, unauthorized access or disclosure, improper disposal, hacking or IT incident, and unknown or other breaches. In the Healthcare Information and Management Systems Society 2019 Cybersecurity Survey, more than 80% of responding hospitals have reported that they experienced a significant security incident in the past 12 months [2].

XSL•FO
RenderX

Another growing cybersecurity threat to hospitals is ransomware attacks. Ransomware denies users the access to data by encrypting the data with a key known only to the attacker [3]. The attacker demands a ransom payment in exchange for the key to decrypt the user's data. In one recent case, a hospital was forced to pay US $17,000 to regain access to its system. California-based Hollywood Presbyterian Medical Center reportedly experienced a malware attack, and employees stated that they were unable to access certain parts of the hospital network [4]. In a more severe case, University of California San Francisco paid over US $1 million to hackers to regain access to its system [5].

The Health Information Technology for Economic and Clinical Health Act regulates the notification of health information breaches in the United States. This act requires health care providers and entities covered by the Health Insurance Portability and Accountability Act of 1996 to notify a breach of protected health information, which affects more than 500 individuals to those affected individuals, HHS, and sometimes the media [1]. HHS maintains a public database called Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information, which publishes the reported health data breaches submitted from October 2009 to the present [6].

Recovering from data breaches and ransomware attacks is costly for hospitals. Data breach remediation efforts were associated with lower hospital quality, including increased time-to-electrocardiogram and an increased 30-day acute myocardial infarction mortality rate [7]. In 2019, the average total cost of a data breach for all industries globally was US $3.92 million, and it took organizations an average of 279 days to identify and contain a breach. The average total cost of a data breach for all industries in the United States was US $8.19 million, which was more than 2-fold the global average [8]. The total costs include notification costs, productivity losses, re-establishing the image of the company, infrastructure costs, and repetition of work. The cost of a data breach is different across industries. The actual cost per breached record averages out at US $242 per record in the United States, and US $150 globally [8]. In the US health care industry, per-record breaches cost an average of US $429 [8]. Global losses from security breaches are forecasted to double from US $3 trillion per year in 2015 to US $6 trillion per year in 2021 [9]. In addition, breached hospitals potentially face investigation, fines, and several years of monitoring by the Office for Civil Rights (OCR) [10].

The additional costs associated with data breaches and their remediation has adverse implications for hospital productivity. The productivity of a firm is typically measured as the value of goods and services produced per unit of labor and capital input. For hospitals, productivity is the value of health care goods and services, such as pharmaceuticals and surgeries, per health care input [11-14]. The financial costs and regulatory burden associated with a data breach may divert resources away from patient care, thus negatively affecting hospital productivity. Disruptions in health IT systems after a breach may disrupt or delay the workflow of clinicians [7], thus negatively affecting hospital productivity. Employee layoffs and turnovers resulting from a breach are another factor that may reduce productivity [15]. Breach remediation required by the OCR, including changes to the health IT system and staff training, may take years to complete. Such an oversight by the OCR, which changes hospital policies and processes may disrupt hospital productivity in the long term. Organizational culture set by hospital administrators may have a strong influence on the productivity and security practices of the staff. Thus, hospitals with poor organizational culture may be involved in a breach and have poor productivity.

Despite the increasing importance of cybersecurity, little is known about its effects on hospital-level productivity. Health IT systems are intended to improve hospital productivity by reducing human error, but data breaches may have the unintended consequence of disrupting hospital productivity. Thus, in this study, we aimed to investigate the relationship between data breaches and hospital productivity by using data from California hospitals from 2012 to 2016. We hypothesized that data breaches may increase hospital costs and disrupt provider workflow, thus decreasing hospital productivity. We compared the productivity of the hospitals that experienced a data breach against control hospitals and investigated whether hospital productivity was significantly different for the breached hospitals before and after a breach.

Methods

Empirical Model

After a data breach, the resulting regulatory enforcement and remediation is a shock to a hospital's patient care delivery. Therefore, hospital data breaches can be modeled as a natural experiment to understand the relationship between data breaches and productivity. The association between hospital data breaches and productivity was estimated using a generalized difference-in-differences model with multiple prebreach and postbreach periods [16]. This model for an event study is a widely used approach to model observational data in the health economics literature.

We used the reported information on breaches as collected by HHS to create a panel of hospital-year observations from 2012 to 2016. Our model estimates the changes in productivity associated with hospitals that experienced a breach, controlling for hospital financial characteristics including total assets, total labor, IT capital, IT labor, bed size, and time trends. The model assumes that the breached hospitals would have followed a trend parallel to that of the control group if they had not been breached.

For a hospital in a given year, the dependent variable is the log of productivity measured as value added. Value added is defined as operating revenues' lesser intermediate inputs. Intermediate inputs include surgical supplies, linens, clothing, and other material inputs [11]. Financial control variables included the log of total capital, total labor, IT capital, and IT labor. Total capital assets include current assets, property, plant and equipment, intangible assets, assets whose use is limited, and other assets. Total labor (non-IT) is defined as the total conventional salaries, wages, employee benefits, and

XSL•FO
RenderX

professional fees excluding any costs related to IT labor. IT capital is a summation of four components: purchased services, leases and rentals, other direct expenditure, and physical capital. IT labor is the summation of salaries and wages, employee benefits, and professional fees associated with data processing. For hospital control variables, we included the number of licensed beds and case mix index of a given hospital. For breach control variables, we included breach type and breach location. In addition, ownership, teaching status, and rural status were included in the descriptive summary, but they were omitted from fixed effects regression because they were time-invariant variables. Finally, the model included year fixed effects and hospital fixed effects. Assuming that hospitals' administration does not change in the short term, hospital fixed effects serve to control for the unobserved time-invariant hospital organizational culture that may be correlated with both breaches and productivity.

For the treatment, dummy hospitals were categorized into two groups: never breached (control) and breached. Moreover, the breached hospitals experienced their specific breach events at different timepoints. The difference-in-differences model was specified to capture changes in value added at –3, –2 –1, 0, +1, +2, and +3 years relative to the hospital-specific year of the data breach. The year of the breach was set as the reference category. For example, a hospital that was breached in 2014 was coded as –2 in 2012, –1 in 2013, +1 in 2013, and +2 in 2014. The coefficients on the event time dummies captured the changes associated with value added at a given timepoint.

The model assumed that a breach was a one-time event. Multiple breaches within a year are a possibility, but we did not find any hospitals that experienced multiple breaches in our sample. The difference-in-differences model was estimated using fixed effects regression. SEs were robust to heteroskedasticity and allowed for within-hospital correlation analysis. Statistical analysis was performed using Stata (version 15, StataCorp) [17].

Data

Breach data and California Hospital financial data were utilized in this study. Breach data published by HHS were used to identify hospital data breaches by hospital name and the date of the breach report [6]. All types of breaches were included (ie, theft, unauthorized access or disclosure, hacking or IT incident, improper disposal, and loss). Only breaches affecting 500 or more individuals were observed in our data; therefore, HHS data do not provide an exhaustive list of all hospital data breaches.

The California Office of Statewide Health Planning and Development (OSHPD) publishes audited financial data from approximately 450 participating nonfederal hospitals licensed by the state. Financial disclosure reports are filed annually by each licensed hospital. OSHPD data provided hospital characteristics and financial variables [18]. Hospital data breaches in the HHS data were merged with OSHPD hospital financial data in accordance with the hospital name and year. OSHPD provides a directory of hospitals and their business names and aliases, which uniquely identify each hospital. However, the HHS data do not provide a standard hospital identifier; thus, some breaches may have been merged incorrectly.

The study sample included general acute care hospitals from 2012 to 2016. For data consistency, hospitals whose financial statements spanned less than 1 year were excluded from the study. Breach activity prior to our study period could influence the response period assessed herein. Thus, hospitals that experienced a breach in the 2 years before our study period (2010 and 2011), were excluded for data consistency. Furthermore, all financial variables were trimmed at the top 1% to exclude outliers. The resulting study sample was an unbalanced panel of hospitals with 2610 unique hospital-year observations. Data breaches were reported by 31 hospital-year observations. The breached group had 205 hospital-years, and the control group had 2405 hospital-years.

Results

Descriptive Statistics

Descriptive statistics are summarized by breach status in Table 1. Hospital year observations were categorized as breached and never breached (control) groups. The number of hospital years was 205 in the breached group and 2405 in the never breached group. The breached group was larger with, on average, more than 2-fold the value added compared to the control group (US $429.4 million vs US $189.55 million, respectively). The breached group had almost 3-fold the total assets (US $685.06 million vs US $254.45 million, respectively) and more than 2-fold the labor spending (US $387.17 million vs US $169.84 million, respectively) than the control group. The breached group spent almost 3-fold more on health IT capital (US $32.43 million vs US $10.83 million, respectively) and spent almost 4-fold more on health IT labor (US $8.54 million vs US $2.20 million, respectively). The breached hospitals were more likely to be larger in bed size (348.8 vs 225.4, respectively) and higher in the case mix index (1.32 vs 1.27, respectively), less likely to be not-for-profit hospitals (43.41% vs 63.49%, respectively), and more likely to be public hospitals (26.34% vs 13.68%, respectively) and teaching hospitals (60.98% vs 6.65%, respectively).

XSL•FO
RenderX

Table 1. Descriptive summary of breached and never breached (control) hospitals.

Variables	Breached (n=205)	Never breached (n=2405)
Continuous variables: financial variables in US $ (million), mean (SD)		
Value-added operating revenue	429.40 (507.97)	189.55 (173.54)
Total assets	685.06 (916.47)	254.45 (323.21)
Total labor	387.17 (413.95)	169.84 (148.41)
Information technology capital	32.43 (71.58)	10.83 (19.07)
Information technology labor	8.54 (15.69)	2.20 (4.12)
Licensed beds	348.80 (211.11)	225.40 (158.10)
Case mix index	1.32 (0.38)	1.27 (0.36)
Categorical variables: ownership, n (%)		
Investor-owned hospitals	62 (30.24)	549 (22.83)
Not-for-profit hospitals	89 (43.41)	1527 (63.49)
Public hospitals	54 (26.34)	329 (13.68)
Teaching hospitals	125 (60.98)	160 (6.65)

A comparison of the financial characteristics of breached and control hospitals between 2012 and 2016 is shown in Table 2. The breached group had a higher growth rate of value added, total assets, and total labor than the control group between 2012 and 2016 (128.27% vs 115.81% for value added, 128.38% vs 121.35% for total assets, and 117.24% vs 111.43% for total labor, respectively). The breached group had a higher growth rate than the control group in IT capital (186.69% vs 178.96%, respectively) and in IT labor (183.96% vs 123.82%, respectively) from 2012 to 2016. The breached group had a higher growth rate in licensed beds (100.39% vs 98.73%, respectively) between 2012 and 2016.

Individuals affected by a breach, breach type, and breach location among breached hospitals are summarized as follows. The mean number of individuals affected by a breach was 136,613. The proportion of breach types indicated that data theft was the most common breach type (65.85%), followed by unauthorized access, loss, or other breach types (22.00%), further followed by hacking or IT incidents (11.71%). The proportion of breach location indicated that desktop computers or laptops were the most common breach locations (51.22%), followed by network servers, papers, films, or other sources (36.1%), further followed by electronic medical records (12.68%).

Table 2. Descriptive summary of breached and never breached (control) hospitals between 2012 and 2016.

Variables	Breached (n=205)		Never breached (n=2405)	
	Mean (SD)	2016 vs 2012, %	Mean (SD)	2016 vs 2012, %
2012, US $ (million)				
Value added	422.06 (494.78)	128.27	193.19 (172.91)	115.81
Total assets	659.66 (870.55)	128.38	262.89 (316.28)	121.35
Total labor	388.65 (416.00)	117.24	174.22 (151.02)	111.43
Information technology capital	29 (71.28)	186.69	9.71 (13.33)	178.97
Information technology labor	7.23 (11.52)	183.96	2.19 (3.33)	123.83
Licensed beds	346.30 (211.13)	100.39	226.96 (160.26)	98.73
2016, US $ (million)				
Value added	541.38 (634.86)	N/A[a]	223.73 (206.48)	N/A
Total assets	846.88 (114.84)	N/A	319.02 (437.35)	N/A
Total labor	455.64 (485.39)	N/A	194.15 (166.60)	N/A
Information technology capital	54.14 (98.85)	N/A	17.37 (36.87)	N/A
Information technology labor	13.30 (23.69)	N/A	2.72 (6.30)	N/A
Licensed beds	347.65 (207.02)	N/A	224.08 (153.52)	N/A

[a]N/A: not applicable.

XSL•FO
RenderX

Regression Results

We estimated the change in value added associated with the years before and after a breach while controlling for hospital assets, labor, IT assets, IT labor, number of beds, case mix index, breach type, breach location, time trends, and hospital fixed effects. The regression coefficients are listed in Table 3 and visualized in Figure 1. We found that productivity remained practically unchanged before and after a breach relative to baseline, with constant observable time-varying covariates, time trends, and hospital fixed effects. Log-transformation of the dependent variable yielded regression coefficients that can be interpreted as multiplicative changes after exponentiation. Specifically, value added was associated with a 0.5% reduction [exp(-0.005)=0.995; *P*=.78] at 1 year after a breach, but the change was not significant. Furthermore, value added was associated with a 1.7% increase [exp(0.017)=1.017; *P*=.32] at 2 years after a breach, but the change was not significant. Moreover, value added was associated with a 2.5% increase [exp(0.025)=1.025; *P*=.28] at 3 years after a breach, but the change was not significant.

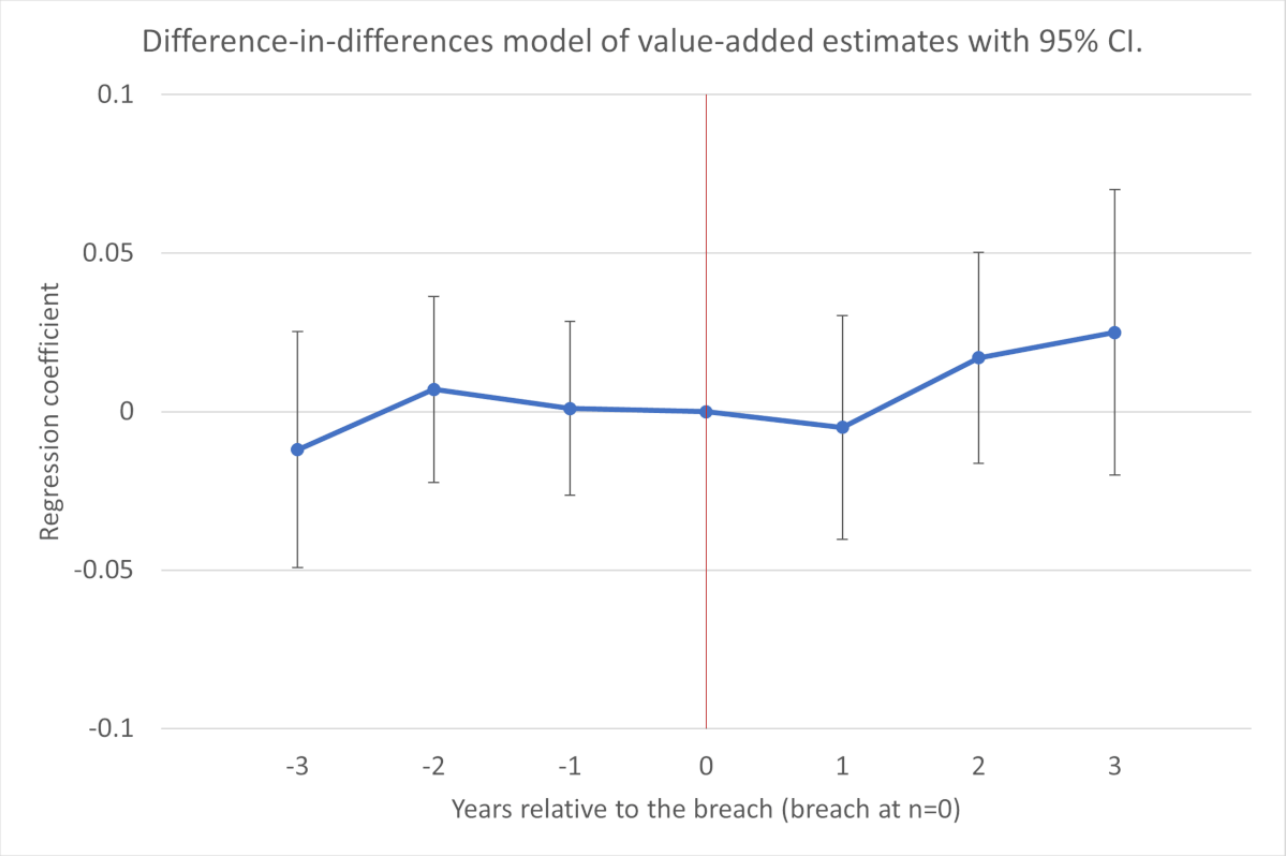
Table 3. Difference-in-differences model estimates for value added.

Breach parameters	Coefficient (SE)	*P* value
Breach time for which ln (revenue) was calculated (reference=0)		
–3	–0.012 (0.019)	.53
–2	0.007 (0.015)	.64
–1	0.001 (0.014)	.94
1	–0.005 (0.018)	.78
2	0.017 (0.017)	.32
3	0.025 (0.023)	.28
Total assets	0.055 (0.016)	.001
Total labor	0.600 (0.064)	<.001
Information technology capital	0.045 (0.007)	<.001
Information technology labor	0.007 (0.003)	.02
Number of beds	0.091 (0.043)	.04
Individuals affected	0.000 (0.000)	.27
Case mix index	0.126 (0.079)	.11
Breach type for which ln (revenue) was calculated		
Hacking or information technology incident (reference)	N/A[a]	N/A
Data theft	0.148 (0.035)	<.001
Unauthorized access, loss, or other	0.108 (0.021)	<.001
Breach location for which ln (revenue) was calculated		
Desktop computer or laptop (reference)	N/A	N/A
Electronic medical record	0.070 (0.033)	.04
Network server, papers, films, or others	0.099 (0.020)	<.001
Year for which the ln (revenue) was calculated (reference=2008)		
2009	0.024 (0.012)	.04
2010	0.042 (0.013)	.001
2011	0.049 (0.014)	.001
2012	0.052 (0.017)	.003
2013	0.037 (0.017)	.03
2014	0.021 (0.017)	.22
2015	0.084 (0.020)	<.001
2016	0.080 (0.024)	.001
Constant	5.042 (1.077)	<.001

[a]N/A: not applicable.

XSL•FO
RenderX

Figure 1. Difference-in-differences model of value-added estimates with 95% CI.



Estimates for all timepoints, from 3 years before to 3 years after a breach, were not significant. These estimates suggest that breaches were not associated with value added.

Total assets, total labor, IT capital, IT labor, and the number of beds were positively associated with value added. The number of individuals affected and the case mix index were not associated with value added. Breach type and breach location were associated with value added.

Discussion

Principal Findings

Hospitals' breach responses increase the financial burden on hospitals. The efforts to repair the damages from a data breach increase direct and indirect costs and may divert resources from improving patient quality of care. Health care data breaches reported to HHS, which includes breached health plans, physicians, and business associates in addition with hospitals, have grown from 329 in 2016 to 642 in 2020 [19]. Hospital data breaches were reported to increase hospital advertising expenditures [20] and IT spending [21] to remedy the damage due to a data breach.

Breached hospitals were larger in size, reflected in higher value added, total assets, and total labor, which is consistent with previous findings [7,22]. Larger hospitals have more access points, devices, and staff that could be breached, both intentionally and erroneously. Thus, the risk of a data breach is proportional to an organization's size.

However, data breaches were not associated with a reduction in productivity; that is, we did not observe a significant relationship between breaches and hospital productivity measured as the value added. Hospital productivity was resilient against the shocks from a data breach. We hypothesized that the financial cost and disruption associated with data breaches may decrease hospital production, but our results suggest that hospital productivity was unaffected. The stability in hospital productivity also implies that patient demand for hospital services was inelastic to data breaches. The remediation efforts and advertising to repair the reputation of the breached hospitals may have contributed to the steady demand.

Moreover, there are at least 2 more reasons to explain these results. First, there is incredible heterogeneity in the information type from a breach. For example, the release of patient records is likely to undermine the reputation of a hospital, whereas malware attacks are more likely to reduce cash flow rather than the hospital's reputation. The effects of different attack types may take longer to manifest for hospitals. Second, while many breaches take place without knowledge, as reflected by the large uncertainty about hospital vulnerabilities, those that detect incidents may not have an incentive to report the full financial impact [23]. Most hospitals are not-for-profit organizations. We are not aware of a federal or state law that requires not-for-profit organizations to disclose data breaches in their financial statements. The Sarbanes–Oxley Act of 2002 requires publicly traded firms to disclose data breaches, but investor-owned hospitals account for a small fraction of all hospitals.

Emphasis should be laid on the security training of health care workers. Treating patients and saving lives are the highest

XSL•FO
RenderX

priority for health care workers, which makes them cautious in handling hospitals' security regulations and policies. However, nearly one-third of the health care workforce had never received cybersecurity-related training [24]. This lack of awareness results in improper handling and storage of patient files, with increasing usage of mobile devices. The most frequent breach type in our study sample was data theft, and the most frequent breach location was desktop and laptop computers. In health care, internal human error and misuse occur much more frequently than external attacks such as those that involve hacking [25]. Thus, to reduce the risk of a hospital data breach, health care workers should be trained in cybersecurity.

Hospitals are an attractive target for cyber attackers, and these attackers are affecting hospitals by using ransomware [26,27]. While our study data do did not capture ransomware attacks, these are considered much more disruptive than data breaches. To mitigate the threat, health care organizations should share threat information, experiences, and best practices to build the appropriate security architecture.

Limitations

Our analysis included reported health data breaches, which affected more than 500 individuals from 2012 to 2016; however, this is not an exhaustive list of data breaches. Smaller data breaches that affect fewer than 500 individuals are not published by HHS; hence, such breaches were excluded from our study. There is a nontrivial number of unpublished small data breaches [28]; however, such breaches tend to be less costly for organizations to remediate. There are various types of data breaches, and given the heterogeneity in potential breach effects, our small sample of breached hospitals limited the precision of our model estimates.

Conclusions

Hospital productivity was resilient against the shocks from a data breach between 2012 and 2016. The productivity trend of breached hospitals remained parallel with that of control hospitals in the years before the breach. Thereafter, the productivity of breached hospitals did not diverge significantly in the years after the breach. Nonetheless, data breaches continue to threaten hospitals today; therefore, health care workers should be trained in cybersecurity to mitigate these disruptions.

Conflicts of Interest

None declared.

References

1. Breach Notification Rule. U.S. Department of Health & Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html [accessed 2020-09-14]
2. HIMSS Healthcare Cybersecurity Survey. Healthcare Information and Management Systems Society, Inc. 2020 Nov 16. URL: https://www.himss.org/himss-cybersecurity-survey [accessed 2020-07-13]
3. Cyber Security Guidance Material. U.S. Department of Health & Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/security/guidance/cybersecurity/index.html [accessed 2021-03-26]
4. Winton R. Hollywood hospital pays $17,000 in bitcoin to hackers; FBI investigating Internet. Los Angeles Times. 2016 Feb 18. URL: https://www.latimes.com/business/technology/la-me-ln-hollywood-hospital-bitcoin-20160217-story.html [accessed 2020-12-29]
5. Drees J. UCSF pays $1M+ ransom to unlock medical school's computer systems. Becker's Health IT. 2020 Jun 29. URL: https://www.beckershospitalreview.com/cybersecurity/ucsf-pays-1m-ransom-to-unlock-medical-school-s-computer-systems.html [accessed 2020-09-14]
6. Cases Currently Under Investigation. U.S. Department of Health and Human Services: Office for Civil Rights. URL: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf [accessed 2020-09-14]
7. Choi SJ, Johnson ME, Lehmann CU. Data breach remediation efforts and their implications for hospital quality. Health Serv Res 2019 Oct;54(5):971-980 [FREE Full text] [doi: 10.1111/1475-6773.13203] [Medline: 31506956]
8. Ponemon L. What's New in the 2019 Cost of a Data Breach Report. SecurityIntelligence. 2019 Jul 23. URL: https://securityintelligence.com/posts/whats-new-in-the-2019-cost-of-a-data-breach-report/ [accessed 2020-07-13]
9. Carole J. Cybersecurity vs Productivity: The CISO's Dilemma. HP Threat Research Blog. 2017 Oct 23. URL: https://threatresearch.ext.hp.com/cybersecurity-vs-productivity-the-cisos-dilemma/ [accessed 2020-09-14]
10. HIPAA Enforcement. U.S. Department of Health & Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/index.html [accessed 2020-09-14]
11. Lee J, McCullough J, Town R. The impact of health information technology on hospital productivity. Rand J Econ 2013;44(3):568. [doi: 10.1111/1756-2171.12030]
12. Sheiner L, Hutchins AM. Measuring productivity in healthcare: an analysis of the literature. Semantic Scholar. 2016. URL: https://www.semanticscholar.org/paper/MEASURING-PRODUCTIVITY-IN-HEALTHCARE-%3A-AN-ANALYSIS-Sheiner-Hutchins/13429271e467091768467f327a3355d93e207318 [accessed 2021-06-16]
13. Harper M, Khandrika B, Kinoshita R, Rosenthal S. Nonmanufacturing industry contributions to multifactor productivity, 1987-2006. Mon Labor Rev 2010;133:16 [FREE Full text] [doi: 10.1787/a40c5025-en]

XSL•FO
RenderX

14. Cylus J, Dickensheets B. Hospital multifactor productivity: a presentation and analysis of two methodologies. Health Care Financ Rev 2007;29(2):49-64 [FREE Full text] [Medline: 18435223]
15. Job Done: Nearly Every Third Corporate Data Breach Gets Employees Fired. Kaspersky. 2018 Sep 12. URL: https://www.kaspersky.com/about/press-releases/2018_job-done-nearly-every-third-corporate-data-breach-gets-employees-fired [accessed 2021-03-26]
16. Jacobson L, LaLonde R, Sullivan D. Earnings Losses of Displaced Workers. The American Economic Review 1993;83:685-709. [doi: 10.17848/wp92-11]
17. Stata. URL: https://www.stata.com/ [accessed 2020-09-14]
18. Office of Statewide Health Planning and Development. URL: https://oshpd.ca.gov/data-and-reports/ [accessed 2020-09-14]
19. Healthcare Data Breach Statistics. HIPAA Journal. URL: https://www.hipaajournal.com/healthcare-data-breach-statistics/ [accessed 2020-10-26]
20. Choi SJ, Johnson ME. Understanding the relationship between data breaches and hospital advertising expenditures. Am J Manag Care 2019 Jan 01;25(1):e14-e20 [FREE Full text] [Medline: 30667613]
21. Choi SJ, Johnson ME, Lee J. An event study of data breaches and hospital IT spending. Health Policy and Technology 2020 Sep;9(3):372-378. [doi: 10.1016/j.hlpt.2020.04.008]
22. Bai G, Jiang JX, Flasher R. Hospital Risk of Data Breaches. JAMA Intern Med 2017 Jun 01;177(6):878-880 [FREE Full text] [doi: 10.1001/jamainternmed.2017.0336] [Medline: 28384777]
23. Collins K. Yahoo and Equifax just proved that you can never trust the first number announced in a data breach Internet. Quartz. 2017 Oct 03. URL: https://qz.com/1093399/the-equifax-efx-and-yahoo-hacks-are-further-proof-that-you-should-never-trust-the-first-number-announced-in-a-data-breach/ [accessed 2020-09-14]
24. Cyber Pulse: The State of Cybersecurity in Healthcare – Part Two. Kaspersky. URL: https://media.kasperskydaily.com/wp-content/uploads/sites/85/2019/08/16121510/Kaspersky-Cyber-Pulse-Report-2019_FINAL.pdf [accessed 2020-09-14]
25. 2019 Data Breach Investigations Report. Verizon. URL: https://enterprise.verizon.com/resources/executivebriefs/2019-dbir-executive-brief-emea.pdf [accessed 2020-09-14]
26. Hospitals are hit with 88% of all ransomware attacks. Becker's Health IT. 2016 Jul 27. URL: https://www.beckershospitalreview.com/healthcare-information-technology/hospitals-are-hit-with-88-of-all-ransomware-attacks.html [accessed 2020-09-14]
27. Gallagher R, Bloomberg. Hackers 'without conscience' demand ransom from dozens of hospitals and labs working on coronavirus. Fortune. 2020 Apr 01. URL: https://fortune.com/2020/04/01/hackers-ransomware-hospitals-labs-coronavirus/ [accessed 2020-09-14]
28. Heubusch K. Little breaches: OCR releases first "small breach" data. J AHIMA 2011 Oct;82(10):56-57. [Medline: 22029215]

Abbreviations

HHS: US Department of Health and Human Services
IT: information technology
OCR: Office for Civil Rights
OSHPD: California Office of Statewide Health Planning and Development

Edited by R Kukafka; submitted 30.11.20; peer-reviewed by M Raj, K Fuji, R Lee, J Marquard; comments to author 15.03.21; revised version received 02.04.21; accepted 04.05.21; published 06.07.21.

Please cite as:
Lee J, Choi SJ
Hospital Productivity After Data Breaches: Difference-in-Differences Analysis
J Med Internet Res 2021;23(7):e26157
URL: https://www.jmir.org/2021/7/e26157
doi:10.2196/26157
PMID:34255672

©Jinhyung Lee, Sung J Choi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Jordan Stillbirth and Neonatal Mortality Surveillance (JSANDS) System: Evaluation Study

Yousef Khader[1], SCD; Mohammad Alyahya[1], PhD; Ziad El-Khatib[2], PhD; Anwar Batieha[1], DrPH; Nihaya Al-Sheyab[1], PhD; Khulood Shattnawi[1], PhD

[1]Department of Public Health, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

[2]Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Yousef Khader, SCD
Department of Public Health
Faculty of Medicine
Jordan University of Science and Technology
Irbid
Jordan
Phone: 962 7 968 02040
Email: yskhader@just.edu.jo

Abstract

Background: The Jordan Stillbirth and Neonatal Mortality Surveillance (JSANDS) is an electronic surveillance system that automatically transfers the data on births, stillbirths, and neonatal deaths to the concerned authorities in the Ministry of Health. JSANDS was implemented and tested in 5 maternity hospitals during the period spanning May 2019 through December 2020.

Objective: This study aimed to evaluate the usefulness and performance of JSANDS to register births, stillbirths, and neonatal deaths, and determine their causes. Specifically, this study examined the JSANDS attributes of acceptability, simplicity, flexibility, stability, representativeness, sustainability, penetration, data quality, sensitivity, and adoption.

Methods: An evaluation study was conducted after 18 months of the JSANDS implementation using the Updated Guidelines for Evaluating Public Health Surveillance Systems. The evaluation focused on how well the system operated to meet its purpose and objectives. The indicators assessing the system attributes were scored on a Likert scale. Each indicator and overall attribute percentage score was represented as score rank and interpreted as excellent (score ≥80%), good (score ≥60 and <80%), average (score ≥40 and <60%), and poor (score <40%).

Results: A total of 270 health care professionals participated in this study and evaluated the system performance. The system users rated the usefulness of JSANDS as excellent (percentage score=85.6%). The overall acceptability (percentage score=82.3%), flexibility (percentage score=80.2%), stability (percentage score=80.0%), and representativeness (percentage score=86.6%) were also rated excellent. The overall simplicity was scored good (percentage score=75.4%). All participants were trained on JSANDS and used it in the past 12 months. Of the 270 respondents, 219 (86.2%) reported that they intend to continue using the JSANDS system to register neonatal deaths and stillbirths in the future. All variables in JSANDS had complete data with no missing values.

Conclusions: The performance of JSANDS in registering all stillbirths and neonatal deaths as well as their causes was excellent. Almost all attributes and indicators of JSANDS functionality were rated excellent. JSANDS can be scaled up to cover all maternity hospitals in Jordan. The potential for scaling up the system is very high for many reasons, including its usefulness, simplified stillbirth and neonatal death review tools, and ease of the reporting process.

(J Med Internet Res 2021;23(7):e29143) doi:10.2196/29143

KEYWORDS

stillbirths; neonatal deaths; surveillance; eHealth; electronic health data; electronic surveillance; pediatrics; maternity; mortality; mortality surveillance; Ministry of Health; health data

XSL•FO
RenderX

Introduction

Maternal and Neonatal Services in Jordan

In Jordan, maternal and neonatal services are provided by the public sector (Ministry of Health and Royal Medical Services), private sector, and teaching hospitals. A total of 27 Ministry of Health hospitals provide birth services, with nearby maternal and child health centers providing antenatal and postnatal care. A recent study in Jordan [1] showed that the quality of maternal and newborn care in many Jordanian hospitals is relatively acceptable. Nonetheless, some maternity hospitals have some deficiencies, including shortages of skilled and competent birth attendants; lack of pivotal protocols, policies, and guidelines necessary for optimal care; lack of optimal thorough antenatal care; and lack of necessary supplies, drugs, equipment, and resources during labor and early postnatal care. Moreover, the study reported that some health care professionals are not regularly trained on routine and special care needed for labor and the early postpartum period in both normal and complicated births.

Jordan Stillbirth and Neonatal Mortality Surveillance

A well-functioning perinatal and neonatal (PNN) mortality surveillance supports the delivery of health services by ensuring the production, analysis, dissemination, and use of reliable and timely information on PNN mortality determinants and quality of care [2-5]. Therefore, a comprehensive electronic surveillance system to register and accurately report stillbirths and neonatal deaths and their causes was developed and implemented.

The Jordan Stillbirth and Neonatal Mortality Surveillance (JSANDS) is an electronic surveillance system that automatically transfers the data on births, stillbirths, and neonatal deaths to the concerned authorities in the Ministry of Health. The development of the surveillance system was informed by the findings of stakeholder analysis and formative qualitative and quantitative research among health professionals, local communities, families, and women [4-7]. The main objectives of this system are to measure the burden of PNN mortality, identify their causes, assess trends in these deaths, and identify risk groups. The data collected by JSANDS are useful in guiding the planning, implementation, and evaluation of programs aiming at prevention and reducing PNN deaths; in prioritizing the allocation of health resources; in formulating research hypotheses; and in providing a basis for epidemiologic research. The JSANDS dashboard displays useful information on mortality indicators, including fetal death rate, neonatal mortality rate, perinatal death rate, and causes of stillbirths and neonatal deaths. These indicators are calculated as an overall score and stratified by social and clinical variables. The potential users of the data include policy makers and decision makers at the Ministry of Health, clinicians, researchers, and other health care providers.

Essential Data Reported by JSANDS

JSANDS collects data on demographic, clinical, maternal, and other relevant characteristics for each live born baby, stillbirth, and neonatal death to allow detailed exploration of the causes of perinatal mortality in Jordan, to calculate and report indicators, to make comparisons between organizations, and to provide a clearer insight into the risk factors most commonly associated with stillbirth or neonatal death. In JSANDS, stillbirth is defined as a baby delivered at or after 24 weeks of gestation or having a birthweight of ≥500 grams and showing no signs of life, irrespective of when the death occurred. Stillbirths include antepartum stillbirth (death before the onset of labor) and intrapartum stillbirth (known to be alive at the onset of labor). Neonatal death is defined as a live born infant at ≥ 24 weeks of gestation or with a birthweight of ≥ 500 grams who died before 28 complete days after birth. Perinatal death is defined as fetal death after 24^{+0} completed weeks of gestation and before 7 complete days after birth.

The Procedure of Death Notification and Assigning Cause of Death

JSANDS reports all live births, stillbirths, and neonatal deaths occurring in each hospital. The obstetrician attending the delivery and the midwife are responsible for entering the necessary data for deliveries and live births. The clinician conducting the delivery of a stillbirth has the primary responsibility to fill the form for the stillbirth, assign the cause of death, and write the International Classification of Diseases Tenth Revision (ICD-10) code accordingly. For neonatal deaths, the clinician and the nurse attending the neonatal deaths are responsible for completing the needed information on each death, including cause of death according to the ICD-10. JSANDS uses the ICD-10 codes to provide a common language for reporting and monitoring diseases [8]. The causes of deaths are set out as follows: main disease or condition in the fetus or infant, for which the single most important and main disease or condition of the infant or fetus who has died is entered; other diseases or conditions in the fetus or infant; main maternal disease or condition affecting the fetus or infant; or other maternal diseases or conditions affecting the fetus or infant.

Objectives of the Evaluation

This study aimed to evaluate the usefulness and performance of JSANDS to register births, stillbirths, and neonatal deaths, and to determine their causes. Specifically, this study examined the acceptability, simplicity, flexibility, stability, representativeness, sustainability, penetration, data quality, sensitivity, and adoption of the system.

Methods

Settings

JSANDS was implemented and tested in 5 maternity hospitals during the period spanning May 2019 through December 2020. These 5 selected hospitals from 3 major governorates cover the vast majority of births and deaths that occur in all regions in the north, east, and south of Jordan. In particular, we included one university teaching, referral hospital in Northern Jordan, which receives labor and delivery cases from all regions and suburbs in Northern Jordan. We also included a governmental hospital affiliated with the Ministry of Health in Northern Jordan, which is specialized for maternal-related issues, including births, and also receives maternal deliveries from all sectors in the region. A large, private hospital in the north of

XSL•FO
RenderX

Jordan was also included from several private and governmental sectors in the region. In the northeast region, we included a large specialized hospital for maternal and child health that covers the whole Al-Mafraq governorate. Finally, we selected another specialized referral hospital in the southern region of Jordan which covers the majority of births within the region.

Study Design

An evaluation study was conducted to ensure that PNN deaths, their causes, and their contributing factors, including social determinates, were being reported and monitored efficiently and effectively by JSANDS after 18 months of the implementation. The evaluation focused on how well the system operated to meet its purpose and objectives. We used evaluation guidelines of the Updated Guidelines for Evaluating Public Health Surveillance Systems developed by the Centers for Disease Control and Prevention (CDC) [9]. The evaluation involved an assessment of the usefulness of the system and other system attributes including acceptability, simplicity, flexibility, stability, representativeness, sustainability, penetration, data quality, sensitivity, and adoption "uptake". These attributes were of concern for the hardware and software, standard user interface, standard data format and coding, appropriate quality checks, and adherence to confidentiality and security standards. For the purpose of evaluation, we engaged all JSANDS users in this study including midwives, nurses, and physicians. The users included those who enter the data, assign causes of death, and use the system for monitoring the deaths in the selected hospitals.

A structured self-reported questionnaire was developed and included specific indicators to assess the system attributes. The indicators were developed based on the literature review of previous public health surveillance evaluations. The questions were developed by 3 experts who have experience in the development and evaluation of surveillance systems, and the content validity was established by a group of 5 specialists in neonatology, obstetrics, and epidemiology. The questionnaire was pilot tested on 12 JSANDS users whose responses were not included in the main study. Minimal changes, mostly related to clarity of questions, were made on these questions. The number of used indicators varied according to the attribute measured. Participants were asked to rate the degree to which they agreed with the attributes' specific indicators using a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree), with a higher score indicating a better performance on the studied attribute.

Operation Definitions of the System Attributes

Level of Usefulness

The system's users were requested to rate 9 indicators of usefulness using a 5-point Likert scale. The system was considered useful if it was able to provide estimates of the PNN mortality rate, the causes of deaths, and the social determinants of deaths; report equity indicators; facilitate assessment of the effect of prevention and control programs; improve the delivery of maternal and child services; and stimulate research.

Acceptability

Seven indicators of acceptability were rated by users. The indicators assessed the user's willingness to use the JSANDS system with and without incentives, their intention to continue using JSANDS in the future, and their satisfaction with the system.

Simplicity

The simplicity of the system included both the structure and ease of operation. Eight simplicity indicators were assessed by users. The indicators assessed the ease of use of the data entry interface, comprehensibility of the variables, comprehensibility of the case definitions, the need for frequent training, and the ease in obtaining the data.

Flexibility

Flexibility referred to whether the system can accommodate new health-related events and changes in case definitions and whether it can be easily integrated with other systems. Ten indicators were assessed by users. The indicators referred the possibility of adding new variables with minimum cost and effort, the functionality of the system after funding withdrawal, the ability of the restructuring the system to monitor maternal deaths, the ability of the restructuring the system to monitor congenital anomalies, the ability of the restructuring the system to monitor infant deaths, the possibility of integrating JSANDS with other systems, the ability of the system to accommodate changes in the case definitions, the ability of the system to accommodate ICD revision, the functionality of the system if used in other countries, and the functionality of the system if used in other health sectors.

Stability

Stability was considered to be the ability of the system to properly collect, manage, and provide data without failure. Five indicators were used to assess the stability. The indicators assessed the stability of the system after withdrawal of sponsors' support, availability of resources for maintenance, and availability of trained staff.

Representativeness

Five indicators were used to assess the usability of the system in other hospitals and the performance of the system to register all births, stillbirths, and neonatal deaths that occurred in the selected hospitals during the study period.

Sustainability

Sustainability referred to the extent to which the system could be maintained or institutionalized within the implementing hospitals. Sustainability was assessed as the success of the research team's activities and the attempt to sustain the system and scale it up to other health facilities, as well as leadership support for sustainability.

Adoption "Uptake"

Adoption uptake was considered to be the intention to continue using the system with or without incentives as assessed by the responses of users engaging with the system at the time of the survey.

XSL•FO
RenderX

Penetration

Penetration was defined as the integration of the surveillance system within hospitals. It was evaluated by calculating the proportion of health professionals who used the surveillance system to report births and PNN deaths relative to the number of trained health professionals.

Data Quality

The database of all registered births and deaths during the study period was reviewed and checked for data quality. The percentage of the missing values in the core variables was calculated.

The sensitivity and positive predictive value of the surveillance system were not calculated because autopsies are not routinely performed for perinatal deaths in Jordan and because the number of perinatal deaths occurring in the community is not known.

Adherence to Ethical Standards

Ethical approval for the JSANDS evaluation was obtained from the Institutional Review Board at Jordan University of Science and Technology. A written informed consent was obtained from all JSANDS users who participated in this study. Every effort was made to protect the confidentiality and the identity of participants.

Data Analysis and Interpretation

The indicators assessing the system attributes were scored on a Likert scale from 1 (“strongly disagree”) to 5 (“strongly agree”). For each indicator, the percentage score was calculated as follows: (sum of all respondents’ scores for each indicator/[maximum score of indicator × number of respondents]) ×100

The overall attribute percentage score was calculated as follows: (sum of all respondents’ scores for all indicators/[number of indicators × maximum score of indicator × number of respondents) ×100. Each indicator and overall attribute percentage score were interpreted as a ranked score as follows: excellent (score ≥80%), good (score ≥60 and <80%), average (score ≥40 and <60%), and poor (score <40%).

Results

JSANDS Users’ Characteristics

Of the 270 JSANDS’s users, 254 (94.1%; 13 men and 241 women) health care professionals participated in this study and evaluated the system performance. Their age ranged from 23 to 65 years, with a mean age of 33.2 years (SD 7.0). Among all respondents, 16 (6.3%) were pediatricians, 23 (9.1%) were obstetricians, 126 (49.6%) were pediatric nurses, and 89 (35.0%) were midwives. All participants were trained on the surveillance system.

Level of Usefulness

The system users rated the usefulness of JSANDS as excellent (percentage score=85.6%). Table 1 shows the ratings of the 9 usefulness indicators. All indicators were rated as excellent. The system provides estimates of the overall PNN mortality rate and the causes of death and those according to other social and clinical determinants. Moreover, the system reports data on other maternal and child health indicators, such as cesarean section rate, preterm delivery rate, and low–birth-weight delivery rate.

Table 1. The percentage score and rank of the JSANDS usefulness indicators.

Usefulness indicators	Score, mean (SD)	Percentage score (%)	Rank
The data in the system can be used to estimate the magnitude of stillbirths and neonatal mortality	4.4 (0.7)	87.8	Excellent
The data in the system can be used to monitor the trend of stillbirths and neonatal mortality	4.4 (0.7)	88.3	Excellent
The data in the system can be used to identifying newborns at high risk of dying	4.3 (0.8)	86.1	Excellent
The data can be used for planning the resources for prevention and control neonatal deaths	4.2 (0.8)	85.0	Excellent
The data in the system can be used to update and develop national policy strategy	4.2 (0.8)	83.5	Excellent
The data in the system can be used to assess the impact of any interventions in your hospital	4.3 (0.8)	85.1	Excellent
JSANDS[a] is able to detect unusual increases in the number of stillbirths and deaths	4.3 (0.7)	86.9	Excellent
JSANDS encourages health workers to prevent neonatal deaths	4.1 (1.0)	82.3	Excellent
JSANDS encourages health workers to improve neonatal health services	4.3 (0.8)	85.7	Excellent
Overall usefulness	4.3 (0.8)	85.6	Excellent

[a]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

The System Performance

Acceptability

The overall acceptability was scored excellent (percentage score=82.3%). All indicators of the acceptability were rated excellent (Table 2) except for 1 indicator, “Using JSANDS is time-consuming,” which received the lowest percentage score (71.8%) with a “good” rank.

XSL•FO
RenderX

Table 2. The percentage score and rank of the JSANDS acceptability indicators.

Acceptability indicators	Score, mean (SD)	Percentage score (%)	Rank
I am very willing to use the JSANDS[a] system and register neonatal deaths and stillbirths	4.3 (0.9)	86.1	Excellent
I am very willing to use the system and register neonatal deaths and stillbirths without having any incentives	4.2 (1.2)	83.1	Excellent
I will continue using the JSANDS system to register neonatal deaths and stillbirths in future	4.2 (0.9)	83.5	Excellent
I am satisfied with the system	4.1 (0.9)	82.1	Excellent
My colleagues are willing to use JSANDS	4.3 (0.8)	86.2	Excellent
The JSANDS team respond to our suggestions and recommendations	4.1 (0.9)	81.8	Excellent
Using JSANDS is not time-consuming	3.6 (1.5)	71.8	Good
Overall acceptability	4.1 (1.0)	82.3	Excellent

[a]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

Simplicity

The overall simplicity was scored as good (percentage score=75.4%). Of the 8 simplicity indicators, 3 indicators were rated excellent and the remaining 5 indicators were rated good (Table 3). The 2 indicators, "The focal points don't need continuous and frequent trainings to use the system" and "The JSANDS doesn't require too much information to register neonatal death and stillbirths," were rated the lowest.

Table 3. The percentage score and rank of the JSANDS acceptability indicators.

Simplicity indicators	Score, mean (SD)	Percentage score (%)	Rank
The JSANDS[a] data entry interface fields are easy to fill	4.4 (0.9)	87.2	Excellent
The variables in the JSANDS data entry interface are easy to understand	4.3 (0.9)	86.5	Excellent
Data entry in the JSANDS is time-consuming	3.6 (1.5)	71.2	Good
Collecting detailed information of cases don't require telephone contact or home visit	3.4 (1.5)	68.7	Good
The case definitions (definition of stillbirth and neonatal death) used by JSANDS are easy to understand	4.1 (1.0)	82.1	Excellent
The focal points don't need continuous and frequent trainings to use the system	3.3 (1.4)	65.9	Good
The sources of information on mortality and causes of deaths are easy to obtain	3.9 (1.3)	77.4	Good
The JSANDS doesn't require too much information to register neonatal death and stillbirths	3.3 (1.5)	65.5	Good
Overall simplicity	3.8 (1.3)	75.4	Good

[a]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

Flexibility

The ratings of the flexibility indicators are shown in Table 4. Overall, flexibility was scored excellent (percentage score=80.2%) by the system users. Seven out of ten flexibility indicators were rated excellent, and the rest were rated good (Table 4).

XSL•FO
RenderX

Table 4. The percentage score and rank of the JSANDS flexibility indicators.

Flexibility indicators	Mean score SD	Percentage score (%)	Rank
New variables can be added to the system with minimum cost and effort	3.7 (1.1)	74.7	Good
The JSANDS[a] system does not need continuous funding	3.9 (1.4)	77.7	Good
The system can be restructured easily to register maternal deaths	4.0 (0.9)	80.3	Excellent
The system can be restructured to register congenital anomalies	4.2 (0.8)	84.0	Excellent
The system can be restructured easily to register infant mortality (deaths below 1 year)	4.1 (0.9)	82.7	Excellent
The system can be integrated in the Hakeem system or your health information system easily	3.8 (1.2)	75.9	Good
The JSANDS system can accommodate changes in case definition of neonatal deaths	4.1 (0.8)	82.4	Excellent
The JSANDS system can accommodate new ICD[b] revisions, such as the ICD-11	4.0 (0.8)	80.0	Excellent
The JSANDS system can be used in any country with minimal changes	4.1 (0.9)	81.9	Excellent
The JSANDS system can be used in any hospital in Jordan	4.1 (1.0)	82.8	Excellent
Overall flexibility	4.0 (1.0)	80.2	Excellent

[a]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

[b]ICD: International Classification of Diseases.

Stability

The overall stability score was rated excellent (percentage score=80.0%) (Table 5). Three indicators were rated excellent, and two indicators were rated good. The indicator, “The system is stable even when sponsors withdrawal support” received the lowest score (72.0%).

Table 5. The percentage score and rank of the JSANDS stability indicators.

Stability indicators	Score, mean (SD)	Percentage score (%)	Rank
There are planned resources for maintenance of the JSANDS[a] system	4.1 (1.0)	82.5	Excellent
The system is stable even sponsors withdrawal support	3.6 (1.4)	72.0	Good
There are trained staff in the hospital to use the system	4.5 (0.6)	90.2	Excellent
It is easy to know how to use the system	4.5 (0.7)	90.4	Excellent
JSANDS is not affected by the turnover of trained staff	4.0 (1.4)	79.8	Good
Overall stability	4.1 (1.0)	80.8	Excellent

[a]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

Representativeness

The overall representativeness score was rated excellent (percentage score=86.6%) (Table 6). All representativeness indicators were rated excellent.

Table 6. The percentage score and rank of the JSANDS representativeness indicators.

Representativeness indicators	Score, mean (SD)	Percentage score (%)	Rank
The system can be used in any MoH[a] hospital	4.2 (0.8)	84.8	Excellent
The system can be used in any Royal Medical services hospital	4.3 (0.8)	85.0	Excellent
The JSANDS[b] system registered all births occurred in the hospital in the past period	4.3 (0.8)	86.6	Excellent
The JSANDS system registered all stillbirths occurred in the hospital in the past period	4.4 (0.7)	88.1	Excellent
The JSANDS system registered all neonatal deaths that occurred in the hospital in the past period	4.4 (0.7)	88.9	Excellent
Overall representativeness	4.3 (0.7)	86.6	Excellent

[a]MoH: Minsitry of Health.

[b]JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance.

XSL•FO
RenderX

Sustainability

The policymakers at the Jordan Ministry of Health and the UNICEF (The United Nations Children's Fund) team were supportive of scaling up JSANDS to other hospitals in Jordan. The Secretory General formed a steering committee for scaling up the system, and a letter signed by the Minister of Health was sent to the other hospitals' directors to implement JSANDS and use it for registration of neonatal deaths and stillbirths.

Adoption Uptake

All participants were trained on JSANDS and used it in the past 12 months. Of the 270 respondents, 219 (86.2%) reported that they intend to continue using the JSANDS system to register neonatal deaths and stillbirths in the future, and 213 (83.9%) reported that they intend to use it even without having any incentives.

Penetration

Of the 270 participants, 14 (5.5%) users used it once, 70 (27.6%) used it but not frequently, 81 (31.9%) used it at least once a weak, and 89 (35.0%) used it every day.

Data Quality and Sensitivity

In Jordan, all births and deaths should be registered and reported to the Civil Status and Passports Department (CSPD). To investigate data quality and sensitivity of the JSANDS, we compared registered deaths on the system with deaths reported to the CSPD for the same period. Further, all data entered on the system were carefully checked by comparing them with data in maternal or child medical records. We found that all variables had complete data with no missing values. Unsurprisingly, all neonatal deaths reported to the CSPD (that used paper notification) from the selected hospitals during the same pilot period were registered in JSANDS. On the other hand, 85.0% (648/762) of neonatal deaths and all stillbirths registered by JSANDS were not registered in the CSPD, neither as births nor as deaths.

Discussion

Counting deaths, collecting information about causes of these deaths, and determining the contributing and avoidable factors are essential for delaying unnecessary early death events and improving the health care system [10]. Meanwhile, designing electronic registration and reporting systems for births and neonatal deaths is essential for reducing stillbirths and neonatal mortality [11].

This study showed that the users of the JSANDS found the system to be useful, functional, and high-performing. In a previous study on the usability of JSANDS, health professionals rated it positively as indicated by the high usability ratings, satisfaction with the system, and ratings of the helpfulness of the system in making important decisions with quality of care [12].

JSANDS was rated by the system users as useful because it provides estimates of the PNN mortality rate and the causes of deaths overall according to other social and clinical determinants. This finding is supported by the fact that the JSANDS data were used by authors in previous studies [13,14] to determine the rate, determinants, and causes of stillbirths and neonatal deaths in Jordan. The researchers in these studies reported that the majority of neonatal deaths could have been prevented and have recommended specialized care to be provided to low–birth-weight neonates and those with respiratory problems by experienced health care providers [13]. Moreover, the researchers reported that the majority of stillbirths registered by JSANDS occurred during the antepartum period, and they recommended that care should be taken for the early identification of high-risk pregnancies and ensuring adequate antenatal obstetric interventions [14].

The system also reports data on other maternal and child health indicators, such as cesarean section rate, preterm delivery rate, and low–birth-weight delivery rate. For example, the system showed that the cesarean section rate is extremely high, reaching a rate of almost 50%. The finding of the high cesarean section rate in Jordan is consistent with the findings of previous studies [15,16].

Reporting mortality estimates by social determinants is useful for informing health care decision making and program evaluation to support the equitable delivery of essential and quality interventions to populations in need. JSANDS collects the data that are necessary to create equity indicators. Combining mortality measures and equity stratifiers in particular ways can yield policy-relevant information that also reveals the basic inequities in society. Having a range of health measures and equity stratifiers facilitates timely recognition of emerging or hidden inequities and improves accountability for protecting vulnerable populations. Currently, the JSANDS dashboard disaggregates mortality measures by gender, nationality, geographic location, and socioeconomic status to clarify if there are any significant differences present.

During the implementation of JSANDS, the availability of data on deaths and their causes triggered health professionals in the hospitals to intervene and improve the quality of services. Several success stories were documented, demonstrating the attempts of health professionals to improve the quality of maternal and neonatal services [17].

JSANDS users rated acceptability and all its indicators as excellent except 1 indicator, "Using JSANDS is time-consuming." One reason for perceiving its use as time-consuming is that JSANDS collects additional information on social determinants of deaths. Including these data is justified as it can improve health equity and reduce disparities in care.

Regarding stability, 3 indicators were rated excellent and 2 indicators were rated good. Although users thought that the stability would be affected by the withdrawal of sponsor support, the JSANDS managers believe that the system will maintain its stability after the end of sponsors' funding because the system is not resource demanding. All hospitals have a sufficient number of computers, and any one of these computers can be used to install JSANDS. The 12 computers that were used for running JSANDS in the pilot hospitals were functional and did not need repair during the period of implementation. The number of unscheduled outages and down times for the system's computers during the 18 months of implementation was 8 times

XSL•FO
RenderX

over all hospitals, and this did not affect the data registration or reporting. The system was operating fully at all times.

The simplicity of JSANDS was rated as good. Five of the simplicity indicators were rated good. It seems that the focal points need continuous and frequent trainings to use the system. To ensure that the focal points are aware of the system, an online manual on JSANDs and its use is available at the JSANDS website [17]. From our own experience in managing JSANDS and training, only 2 hours of training are needed to train people on using the system. Moreover, JSANDS users thought that JSANDS requires too much information to register neonatal death and stillbirths. Looking at an objective indicator of time needed to enter the data into JSANDS, the time spent on collecting and entering the data for registering a birth ranged from 3 to 5 minutes while the time needed for registering a neonatal death and stillbirth was 2 and 3 minutes, respectively. Once data about each case are entered into the system at the health facility, the data are transferred immediately to the Ministry of Health. Moreover, the system automatically analyzes the data and produces the indicators and reports promptly once they are requested.

JSANDS was piloted in 5 hospitals, and the representativeness of deaths cannot be assessed. However, the system was able to accurately describe the occurrence of PNN deaths over time and its distribution in the selected hospitals. On the other hand, users reported that the system can be used in all health sectors. With the use of more objective measures of representativeness, all births, stillbirths, and neonatal deaths that occurred in the selected hospitals were registered and reported by JSANDS.

In conclusion, JSANDS was shown to be a reliable and useful system for registering and reporting stillbirths and neonatal deaths and their causes in Jordan. This real-time surveillance system enables health care professionals to detect unusual increase in stillbirths and neonatal deaths and to detect the exact root of the noticeable deaths at a particular time. The performance of JSANDS in fully registering stillbirths and neonatal deaths and their causes was excellent. Almost all attributes and indicators of JSANDS functionality were rated excellent. JSANDS can be scaled up to cover all maternity hospitals in Jordan. The potential for scaling up the system is very high for many reasons, including its usefulness, the simplified PNN death review tools, and its easy-to-use reporting process.

Acknowledgments

We acknowledge the International Development Research Centre-Canada and UNICEF-Jordan for their financial and technical support.

Conflicts of Interest

None declared.

References

1. Khader Y, Alyahya M, Al-Sheyab N, Shattnawi K, Saqer H, Batieha A. Evaluation of maternal and newborn health services in Jordan. J Multidiscip Healthc 2018;11:439-456 [FREE Full text] [doi: 10.2147/JMDH.S171982] [Medline: 30233201]
2. Batieha A, Khader Y, Berdzuli N, Chua-Oon C, Badran E, Al-Sheyab N, et al. Level, causes and risk factors of neonatal mortality, in Jordan: results of a national prospective study. Matern Child Health J 2016 May;20(5):1061-1071 [FREE Full text] [doi: 10.1007/s10995-015-1892-x] [Medline: 26645614]
3. Obeidat N, Khader Y, Batieha A, Abdel Razeq N, Al-Sheyab N, Khassawneh M. Neonatal mortality in Jordan: secondary analysis of Jordan Population and Family Health Survey (JPFHS) data. J Matern Fetal Neonatal Med 2019 Jan;32(2):217-224 [FREE Full text] [doi: 10.1080/14767058.2017.1377174] [Medline: 28982268]
4. Khader YS, Batieha A, Khader A, Hamadneh S. Stillbirths in Jordan: rate, causes, and preventability. J Matern Fetal Neonatal Med 2020 Apr 25;33(8):1307-1314 [FREE Full text] [doi: 10.1080/14767058.2018.1517326] [Medline: 30153760]
5. Khader Y, Bawadi H, Khasawneh W, Alyahya M, Shattnawi K, Al-Sheyab N, et al. Sociocultural, political, and health system-related determinants of perinatal deaths in Jordan from the perspectives of health care providers: a qualitative study. J Matern Fetal Neonatal Med 2020 Jul 29:1-10 [FREE Full text] [doi: 10.1080/14767058.2020.1800632] [Medline: 32727234]
6. Khader Y, Al-Sheyab N, Alyahya M, Batieha A. Registration, documentation, and auditing of stillbirths and neonatal deaths in Jordan from healthcare professionals' perspectives: reality, challenges and suggestions. J Matern Fetal Neonatal Med 2020 Oct;33(19):3338-3348 [FREE Full text] [doi: 10.1080/14767058.2018.1531120] [Medline: 30348027]
7. Alyahya M, Khader Y, Batieha A, Asad M. The quality of maternal-fetal and newborn care services in Jordan: a qualitative focus group study. BMC Health Serv Res 2019 Jun 26;19(1):425 [FREE Full text] [doi: 10.1186/s12913-019-4232-9] [Medline: 31242940]
8. The WHO application of ICD-10 to deaths during the perinatal period: ICD-PM, 2016. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/249515/9789241549752-eng.pdf?sequence=1 [accessed 2021-06-28]
9. German R, Lee L, Horan J, Milstein R, Pertowski C, Waller MN, Guidelines Working Group Centers for Disease ControlPrevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: 18634202]

XSL•FO
RenderX

10. Bhutta Z, Das J, Bahl R, Lawn J, Salam R, Paul V, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? The Lancet 2014 Jul;384(9940):347-370 [FREE Full text] [doi: 10.1016/s0140-6736(14)60792-3]
11. Lawn J, Blencowe H, Oza S, You D, Lee A, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. The Lancet 2014 Jul;384(9938):189-205 [FREE Full text] [doi: 10.1016/s0140-6736(14)60496-7]
12. Khader Y, Shattnawi K, Al-Sheyab N, Alyahya M, Batieha A. The usability of Jordan stillbirths and neonatal deaths surveillance (JSANDS) system: results of focus group discussions. Arch Public Health 2021 Mar 07;79(1):29 [FREE Full text] [doi: 10.1186/s13690-021-00551-1] [Medline: 33678194]
13. Al-Sheyab N, Khader Y, Shattnawi K, Alyahya M, Batieha A. Rate, risk factors, and causes of neonatal deaths in Jordan: analysis of data from Jordan stillbirth and neonatal surveillance system (JSANDS). Front Public Health 2020;8:595379 [FREE Full text] [doi: 10.3389/fpubh.2020.595379] [Medline: 33194998]
14. Shattnawi K, Khader Y, Alyahya M, Al-Sheyab N, Batieha A. Rate, determinants, and causes of stillbirth in Jordan: Findings from the Jordan Stillbirth and Neonatal Deaths Surveillance (JSANDS) system. BMC Pregnancy Childbirth 2020 Sep 29;20(1):571 [FREE Full text] [doi: 10.1186/s12884-020-03267-2] [Medline: 32993562]
15. Salem A. Trends in Caesarean section deliveries in Jordan from 1982 to 2017: retrospective analyses of annual hospital reports. East Mediterr Health J 2021 Feb 25;27(2):195-201 [FREE Full text] [doi: 10.26719/2021.27.2.195] [Medline: 33665804]
16. Al Rifai R. Rising cesarean deliveries among apparently low-risk mothers at university teaching hospitals in Jordan: analysis of population survey data, 2002-2012. Glob Health Sci Pract 2014 May;2(2):195-209 [FREE Full text] [doi: 10.9745/GHSP-D-14-00027] [Medline: 25276577]
17. Events and news. Jordan Stillbirth and Neonatal Mortality Surveillance System (JSANDS). URL: http://www.jsands.jo/events-and-news/# [accessed 2021-06-28]

Abbreviations

CDC: Centers for Disease Control and Prevention
CSPD: Civil Status and Passports Department
ICD-10: International Classification of Diseases Tenth Revision
JSANDS: Jordan Stillbirth and Neonatal Mortality Surveillance
PNN: perinatal and neonatal
UNICEF: The United Nations Children's Fund

Edited by R Kukafka; submitted 28.03.21; peer-reviewed by F Lami, B Nievas Soriano; comments to author 26.04.21; revised version received 26.04.21; accepted 24.05.21; published 21.07.21.

Please cite as:
Khader Y, Alyahya M, El-Khatib Z, Batieha A, Al-Sheyab N, Shattnawi K
The Jordan Stillbirth and Neonatal Mortality Surveillance (JSANDS) System: Evaluation Study
J Med Internet Res 2021;23(7):e29143
URL: https://www.jmir.org/2021/7/e29143
doi:10.2196/29143
PMID:34287214

©Yousef Khader, Mohammad Alyahya, Ziad El-Khatib, Anwar Batieha, Nihaya Al-Sheyab, Khulood Shattnawi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recruitment and Retention Strategies Among Racial and Ethnic Minorities in Web-Based Intervention Trials: Retrospective Qualitative Analysis

DaSol Amy Hwang[1], MSN, RN, MAT; Alex Lee[1], MSN, RN; Jae Min Song[1], MSN; Hae-Ra Han[1], PhD, RN, FAAN

Johns Hopkins University School of Nursing, Baltimore, MD, United States

Corresponding Author:
Hae-Ra Han, PhD, RN, FAAN
Johns Hopkins University School of Nursing
525 N Wolfe St
Room 533
Baltimore, MD, 21205
United States
Phone: 1 410 614 2669
Email: hhan3@jh.edu

Abstract

Background: Racial and ethnic minority groups are underrepresented in health research, contributing to persistent health disparities in the United States. Identifying effective recruitment and retention strategies among minority groups and their subpopulations is an important research agenda. Web-based intervention approaches are becoming increasingly popular with the ubiquitous use of the internet. However, it is not completely clear which recruitment and retention strategies have been successful in web-based intervention trials targeting racial and ethnic minorities.

Objective: This study aims to describe lessons learned in recruiting and retaining one of the understudied ethnic minority women—Korean Americans—enrolled in a web-based intervention trial and to compare our findings with the strategies reported in relevant published web-based intervention trials.

Methods: Multiple sources of data were used to address the objectives of this study, including the study team's meeting minutes, participant tracking and contact logs, survey reports, and postintervention interviews. In addition, an electronic search involving 2 databases (PubMed and CINAHL) was performed to identify published studies using web-based interventions. Qualitative analysis was then performed to identify common themes addressing recruitment and retention strategies across the trials using web-based intervention modalities.

Results: A total of 9 categories of recruitment and retention strategies emerged: authentic care; accommodation of time, place, and transportation; financial incentives; diversity among the study team; multiple, yet standardized modes of communication; mobilizing existing community relationships with efforts to build trust; prioritizing features of web-based intervention; combined use of web-based and direct recruitment; and self-directed web-based intervention with human support. Although all the studies included in the analysis combined multiple strategies, prioritizing features of web-based intervention or use of human support were particularly relevant for promoting recruitment and retention of racial and ethnic minorities in web-based intervention trials.

Conclusions: The growing prevalence of internet use among racial and ethnic minority populations represents an excellent opportunity to design and deliver intervention programs via the internet. Future research should explore and compare successful recruitment and retention methods among race and ethnic groups for web-based interventions.

Trial Registration: ClinicalTrials.gov NCT03726619; https://clinicaltrials.gov/ct2/show/NCT03726619.

(J Med Internet Res 2021;23(7):e23959) doi:10.2196/23959

KEYWORDS

recruitment and retention; web-based intervention; clinical trial; Korean American; racial/ethnic minority

XSL•FO
RenderX

Introduction

Background

Health research creates generalizable scientific knowledge, leads to positive changes in health policy, and improves human health and health care [1,2]. However, these benefits are not shared equally, particularly among the racial and ethnic minority groups in the United States, as they are vastly underrepresented in biomedical research. For example, racial and ethnic minority participation in cancer clinical trials was approximately 20% compared with 80% of White counterparts in 2018 [3], and this has yielded only a 2% increase in the past two decades. Racial and ethnic minorities constituted 40% of the US population in 2018 [4], and this number is projected to continue to rise, making up more than 50% of the total population by 2050 [5]. The National Institutes of Health policy states that all National Institutes of Health–funded clinical research must include and address minority groups and their subpopulations in their research objectives, design, and recruitment [2]. Therefore, effective recruitment and retention strategies for racial and ethnic minority populations must be identified, evaluated, and used to reduce the underrepresentation of racial and ethnic minorities in health research.

A systematic review of studies addressing research participation among racial and ethnic minorities [6] revealed various factors salient to racial and ethnic minority participation in research. The most common barriers across all 4 ethnic minority groups—Blacks or African Americans, Asian Americans, Latinos, and Pacific Islander—were mistrust (77%), followed by time and financial constraints (45%), fear of unintended outcomes (31%), limited access to information (31%), stigma about participating in research (27%), fear of discrimination related to health insurance coverage (7%), and fear of deportation and concern about their immigration status being affected (5%). In addition, a substantial number of racial and ethnic minority groups experience logistical problems such as transportation, time commitment, and financial means to make office visits [7], lack of social support, low socioeconomic status, lack of culturally targeted or sensitive information, and education about the study [8]. Finally, provider biases against minority patients, such as assumptions about patient preferences, knowledge, and compliance, especially when accompanied with language barriers, may contribute to a provider's failure to offer opportunities for their patients to participate in research [7]. This lack of research participation could hinder racial and ethnic minority individuals from exploring different treatment options and receiving high-quality care while contributing to future generations.

With the rapid advancement of technology and its convenience, access to the internet is highly prevalent. According to the Health Information National Trends Survey by National Cancer Institute in 2007 [9], 70% of 4092 Health Information National Trends Survey respondents indicated that they had access to the internet, excluding work and public spaces, and there was consistent use of social networking sites among racial and ethnic minorities. According to the World Data sourced by the World Bank, the share of the United States population using the internet increased from 69% in 2006 to 76% in 2016 [10]. Similarly, a recent report from the Pew Research Center showed that the internet use penetration rate is equally, if not more rapidly, increasing among Blacks and Hispanics by 17% and 15% between 2010 and 2019, respectively, as compared with that of their White counterparts, which increased by 14% [11]. The growing prevalence of internet use, particularly among racial and ethnic minority populations, represents an excellent opportunity for researchers to design and deliver intervention programs using technology. Indeed, an increasing number of clinical trials have used the internet as the main mode of intervention delivery [12]. However, there is a lack of research addressing recruitment and retention strategies among racial and ethnic minority populations in web-based intervention trials, suggesting an important gap in the literature.

Objective

e-CHECC-uP (e-Community–Based Health Literacy–Focused Intervention for Cancer Control) is a recently completed randomized controlled pilot trial (trial registration: ClinicalTrials.gov NCT03726619) in which a web-based intervention was used to promote breast and cervical cancer screening among Korean American women. The purpose of this paper is to describe our experiences and lessons learned in recruiting and retaining Korean American women for the e-CHECC-uP study. We also compare our findings with the strategies reported in relevant published studies to enhance our understanding of the recruitment and retention of ethnic and racial minorities in web-based intervention trials.

Methods

Description of the e-CHECC-uP Study

e-CHECC-uP was a community-based randomized controlled pilot trial designed to promote mammogram and Papanicolaou test screening practices among Korean American women. Details regarding the study design and intervention are described elsewhere. Briefly, the e-CHECC-uP intervention consisted of web-based health literacy education followed by phone counseling for 6 months. The study inclusion criteria were as follows: (1) Korean American women aged between 21 and 65 years, (2) residing in the Baltimore-Washington metropolitan area, and (3) who were overdue for an age-appropriate mammogram (ie, had not had a mammogram within the past year for women aged 40-54 years or past 2 years for women aged 45-65 years) [13] or a Papanicolaou test (ie, had not had a Papanicolaou test within the last 3 years) [14]. Our recruitment goal was 40 Korean American women (20 in the intervention arm and 20 in the control arm) for this pilot study.

Upon approval of study procedures by the Johns Hopkins Medicine Institutional Review Board, the study team worked with ethnic churches in the target area to identify potential participants by posting flyers, working directly with pastors to make an announcement about the study, and by word-of-mouth among the potential participants. Specifically, after the trained bilingual research assistants identified and contacted the spokesperson (pastor, office secretary, or community health worker) from each church via phone calls and emails, a study coordinator sent an introductory email with a study flyer in

XSL•FO
RenderX

Korean and English and a summary of the study for the church bulletin board announcement. Once approved by the church leadership board, which often included pastors and older adults, in-person baseline visits were scheduled at their convenience, usually during lunchtime or after church service on Sundays. Trained bilingual research assistants met potential participants at their church to explain the study procedure, obtain informed consent, and provide study-related assistance. The enrolled participants completed the study surveys at baseline and at 3 months and 6 months. To address scheduling challenges and time constraints for some potential participants, we also offered an option to complete the consent process and data collection on the web. Every participant received financial incentives at each data collection (US $20 each at baseline and at 3 months and US $40 at 6 months). At the conclusion of the study survey at 6 months, women in the intervention arm were invited to participate in postintervention phone interviews. Each interview lasted for 15-30 minutes. Those who participated in the postintervention interview received additional US $30.

Sources of Data

We used multiple data sources for the e-CHECC-uP trial to address the study purpose, including the study team's weekly meeting minutes, transcripts for postintervention interviews, recruitment and retention tracking logs, survey reports, and other study forms recording contacts with participants such as phone counseling logs. In addition, key lessons learned for recruitment and retention from published studies using web-based interventions were identified and compared with the e-CHECC-uP strategies. Our search was conducted in April 2020, and the strategy involved 2 electronic databases: PubMed and CINAHL. In consultation with a health science librarian, we used the following search terms: *research subject recruitment* OR *research subject retention* OR *recruitment strategies* OR *retention strategies* OR *research participation* AND *web-based* OR *Internet* OR *computers* AND *ethnic groups* OR *ethnic minority*. We included the articles that involved interventions with technology or the internet for racial and ethnic minority populations in the United States, discussed recruitment and retention strategies, and were published in the last 10 years in English. We excluded the studies in which involvement of technology was limited to the recruitment method only. Our search yielded 163 articles, and following the abstract and full-text screening, 7 articles, including 2 systematic reviews, were identified as being relevant to our discussion.

Analysis

Using diverse sources of data, recruitment and retention strategies used in the e-CHECC-uP trial were described, discussed, and solidified among the research team members. Descriptive statistics such as means, frequencies, and percentages were used to summarize recruitment and retention rates in the e-CHECC-uP trial. A thematic analysis of the intervention participant interview was conducted to identify the common experiences of Korean American women with the study intervention in the trial.

Results

Characteristics of e-CHECC-uP and Other Relevant Studies

Table 1 summarizes the key characteristics of the e-CHECC-uP study and the 7 relevant studies that used web-based intervention, as identified from our electronic search. The e-CHECC-uP study participants were recruited over a span of 6 months (July to December 2019) from 5 Korean churches in the Greater Baltimore area, and 72% (52/72) women who were screened for this study were considered as eligible for enrollment. However, 8 of those eligible women refused to enroll in our study because of time constraints (n=4), health reasons (n=1), lack of interest (n=1), and reluctance to undergo a cancer screening test in the United States because of inconvenience (n=1). In addition, we could not reach 4 eligible women after initial contact, yielding 40 women enrolled in the study (77% response rate). The majority of women (n=38) were enrolled face-to-face and 2 were enrolled on the web. In total, 34 participants (intervention: n=15 and control: n=19) completed the final study assessment at 6 months (retention rate 85%). Reasons for withdrawal included time constraints (n=3), participation burden (n=1), change of mind (n=1), and lost contact (n=1). The scope of web-based interventions included in the relevant studies addressed health promotion and participant education. The follow-up period for these studies ranged from 3 weeks to 24 months. Of the 7 relevant studies, 4 randomized controlled trials (RCTs) and 1 RCT-based secondary analysis focused on various racial and ethnic minority groups, including Asian Americans (Chinese, Korean, and Japanese, specifically), African Americans, and Black or Latino. The interventions included the Fitbit device, weekly interactive web-based comics with text and email reminders, web-based education involving in-person visits, and social media websites using a mobile chat function in participants' mother language. There were 2 systematic reviews of e-mental health interventions among Asian American women and mobile health (mHealth) research participation among African Americans, including 79 studies and 56 studies, respectively.

XSL•FO
RenderX

Table 1. Characteristics of studies with web-based interventions involving racial and ethnic minorities.

Reference	Study design; goal	Sample	Follow-up period	Intervention	Recruitment and retention methods with rates
Hwang et al (unpublished, 2021) e-CHECC-uP[a]	RCT[b] (2 arms); to promote mammogram and Papanicolaou test screening among Korean American women	Korean American women who were overdue for an age-appropriate mammogram and/or Papanicolaou test screening (n=40)	6 months	Intervention: web-based health literacy–focused education followed by monthly phone counseling+cancer screening brochure; control: cancer screening brochure	Posting flyers or working directly with pastors to make study announcements in ethnic churches and word-of-mouth; 77% recruitment yield rate; the retention rate was 85% at 6 months
Chee et al (2019) [15]	RCT (3 arms); to promote physical activity among Asian American women	Chinese and Korean women who were aged 20-60 years, who were fluent in English or native language, and who had internet access (n=165)	3 months	Group 1: web-based physical activity promotion; group 2: 1+Fitbit Charge; group 3: 2+Fitbit+office visits	Contacted Asian American groups and communities to announce about the study via email or website; 76% recruitment yield rate. Retention rate was 23% at 3 months (22% for group 1, 37% for group 2, and 0% for group 3)
DeFrank et al (2019) [16]	RCT (2 arms); to decrease obesity risk among urban minority youth	Black or Latino child (aged 9-12 years)-parent dyads, with child BMI percentile of 5% or higher and internet access (n=89)	3 months	Intervention: a 6-chapter interactive web-based nutrition comic (1 chapter per week) combined with weekly text or email messages and reminders from comic characters and self-assessments of how they did on their weekly health goals; control: 6 web-based newsletters with nutrition-related content (1 per week) with weekly text or email messages and self-assessments	Recruitment rates yielded by the various recruitment methods: community events (39%), community tabling or handing out flyers (34%), friends or word-of-mouth (15%), and community clinic referrals (12%); the retention rate was 84% at 3 months
Staffileno et al (2017) [17]	RCT (2 arms); to evaluate the feasibility of web advertising as a recruiting modality for a healthy lifestyle behavior change intervention targeting young African American women	African American women (aged 18-45 years) with untreated prehypertension and internet access (n=35)	12 weeks	Web-based education (12 modules) focusing on either physical activity or Dietary Approaches to Stop Hypertension	Study flyers, tabletop cards, blood pressure screenings at health fairs, and clinic referrals. Web-related methods, such as Facebook, craigslist, university website, intranet, and “on-hold” telephone line, were also used; for 18 months, 176 inquiries received, and 35 women met the criteria and enrolled (20% overall rate) with rates by different recruitment methods used: employee-in-services (100%), university website (44%), flyers or tabletop cards (32%), word-of-mouth or physician referral (25%), blood pressure screenings, (15%), Facebook or Craigslist (13%), and clinics (12%)
Im et al (2020) [18]	RCT (2 arms); to determine the effect of a technology-based education and coaching intervention on improving Asian American breast cancer survivor experience	Asian American breast cancer survivors (n=94)	3 months	Intervention: 3 social media websites with preferred language chat function about breast cancer; web-based education sessions that were culturally tailored (ie, acupuncture for Chinese and red ginseng for Koreans); and web-based cultural resources and the American Cancer Society website for 3 months; control: use of the American Cancer Society website for 3 months	Social media groups, internet communities or groups, and communities or groups of Asian Americans using communication apps that are well known to the subethnic groups (weChat for Chinese, KakaoTalk for Korean, and Line for Japanese); recruitment and retention rates not reported

XSL•FO
RenderX

Reference	Study design; goal	Sample	Follow-up period	Intervention	Recruitment and retention methods with rates
Reyes et al (2018) [19]	Systematic review; to address challenges, barriers, and strategies of conducting e-mental health intervention research among Asian American women	79 studies related to mental health services (not all specific to Asian American women)	No follow-up period discussion provided	No discussion on the scope and characteristics of the interventions included in the review	Participants in the studies included in the review recruited from clinical settings were less likely to complete the program than those recruited from nonclinical settings. Otherwise, no other specific information on recruitment and retention was provided; recruitment and retention rates not discussed
Im et al (2016) [20]	Secondary analysis of an internet-based intervention study using RCT with Asian American breast cancer survivors[c]; to identify practical issues in web-based recruitment of Asian American breast cancer survivors	Asian American (Chinese, Korean, and Japanese) breast cancer survivors (n=57)	3 months	Intervention: used both the internet cancer support group and the internet resources (those related to daily life and those by the American Cancer Society); control: used only the internet resources related to daily life (eg, news in Asian countries and Asian businesses in the United States) and the American Cancer Society website	Study announcement via internet breast cancer support groups, Facebook, and WeChat (reached 8283 potential participants but yielded only 69 study website clicks); 171 internet breast cancer support groups identified, and 6 of them responded to study team email request to post study announcement, with 0.2% response rate of potential participants; retention rates were approximately 50% across the 3 subethnic groups
James et al (2016) [21]	Systematic review; to review participation of African Americans in eHealth or mobile health interventions	Studies that were published in English, conducted in the United States, included African American adults (aged ≥18 years), and specified the type of technology used (n=56)	Ranges from 3 weeks to 24 months	No description of the scope and characteristics of eHealth or mobile health interventions included in the review	Recruitment settings or methods used in the studies included in the review: clinics or health care facilities, 52% (29/56, of studies included); flyer and newspaper, 25% (14/56); websites, listserv, email, or postal mailings, 15% (8/56); churches, n=2; college campus, n=3; no recruitment and retention rates reported

[a]e-CHECC-uP: e-Community–Based Health Literacy–Focused Intervention for Cancer Control.

[b]RCT: randomized controlled trial.

[c]Original study citation not provided.

Recruitment and Retention of Racial and Ethnic Minorities in Web-Based Intervention Trials

Recruitment and Retention in Other Web-Based Intervention Trials

Analysis of various data sources revealed several key strategies and lessons learned relevant to the recruitment and retention of racial and ethnic minorities in web-based intervention trials (Textbox 1). These included authentic care; accommodation of time, place, and transportation; financial incentives; diversity among study team; multiple, yet standardized, modes of communication; mobilizing existing community relationships with efforts to build trust; prioritizing features of web-based intervention; combined use of web-based and direct recruitment; and self-directed web-based intervention with human support.

XSL•FO
RenderX

Textbox 1. Key strategies for the recruitment and retention of racial and ethnic minorities from studies using web-based interventions.

Authentic Care

- Active engagement individualized for participants (eg, assigning study staff to participants and ongoing study process quality assessment) [16,19,21]
- Accommodating participants' family members (eg, separate hospitality room, entertainment, and snacks) [16]
- Continuous showing of genuine care with respect (e-Community–Based Health Literacy–Focused Intervention for Cancer Control [e-CHECC-uP] study)

Accommodation of Time, Place, and Transportation

- Flexible scheduling (eg, 7 days a week, including after school and evenings, weekends) [16], (e-CHECC-uP study)
- Accessible location and providing transportation for in-person data collection in consideration of low-income and minority population [16,21] (e-CHECC-uP study)
- Changing from in-person visits to fully remote work to minimize in-person contact while adapting to recommended health care guidelines during COVID-19 for participants' health and safety (e-CHECC-uP study)

Financial Incentives

- Incentives matching with study participants' interest [16]
- US $50 electronic gift cards [18]
- US $20 gift card at each in-person visit [17]
- Incentives periodically or at the end [21] (e-CHECC-uP study)

Diversity Among Study Team

- Multilingual or bilingual research staff reaching out and delivering the intervention to a population with low socioeconomic status and low English proficiency [16,18,20] (e-CHECC-uP study)
- Study team members with technological proficiency to enhance the participant's learning and use of web-based content (e-CHECC-uP study)

Multiple, Yet Standardized, Modes of Communication

- Multiple modes of communication (eg, phone, email, and text messaging) to provide a diverse platform for efficient interactions between the study team and study participants [16,19] (e-CHECC-uP study)
- Standardized and unified communication system (eg, dedicated study phone line or email address) [16,19]
- Visit reminder protocol via automated text messaging (e-CHECC-uP study)

Mobilizing Existing Community Relationships With Efforts to Build Trust

- Working with community consultants and leaders (eg, listening to a pastor's advice to visit the community site) [15] (e-CHECC-uP study)
- Using a peer advocate group to enhance retention among patients with breast cancer [21]
- Community-based recruitment (eg, targeted flyers posted at local businesses, partnerships with community- or faith-based organizations, and event tabling) [16] (e-CHECC-uP study)
- Dissemination of study progress with sharing of research experiences and findings through writing or voice recording [22] or video recording (e-CHECC-uP study)

Prioritizing Features of the Web-Based Intervention

- Prioritizing features of the web-based intervention (ie, ease and convenience of use, perceived helpfulness, and likelihood of future use) [19]
- Developing thorough guidelines regarding interventions with technology in participants' own languages with precreated IDs and passwords [15] (e-CHECC-uP study)

Combined Use of Web-Based and Direct Recruitment

- Web-based strategies (eg, university website or Craigslist) have greater potential to reach a larger number of people with less financial and human resources but are not as effective as traditional in-person strategies [17]
- A combination of internet setting and physical community settings to recruit target sample size [20] (e-CHECC-uP study)

Self-Directed Web-Based Intervention With Human Support

- Having participants monitor their own progress with goals of building confidence and persistence and solving issues and concerns [18]

XSL•FO
RenderX

- Human support (eg, professionals, peers, and community health workers) to promote participant engagement in the intervention activities [19] (e-CHECC-uP study)

Authentic Care

The Belmont Report issued in 1978 articulates 3 ethical principles that researchers should abide by: respect for persons, beneficence, and justice. Applications in the research setting include informed consent, assessing the participants' comprehension and voluntariness, providing additional protection for those with diminished autonomy, maximizing benefits and minimizing harm to study participants, and fair distribution of research benefits and burdens among diverse populations [23]. As such, authentic care can be translated into the active engagement of study participants in the study process. According to Woolf et al [24], authentic engagement not only enhances the study participants' understanding of data but also ensures their participation as stakeholders and active community agents to further disseminate study findings, develop supplemental resources, and collaborate with local leaders and policy makers. The collective impact of authentic care delivery is rooted in beneficence. Although authentic care may not be measurable by numerical data, our participants in the e-CHECC-uP trial stated their appreciation during the postintervention interview. A participant shared:

> *...I felt like I was being taken care of...I saw how you were trying hard to encourage me and share with me what you know.*

Another participant said:

> *Honestly, going to the hospital takes a toll...I really did not want to go so I kept delaying my appointment, but I felt good after the [cancer screening] visit. Then I thought 'wow, this push was all for me, not for [the research team]...*

The shared experience of our participants shows that authentic care in our communication and interactions was a vital part of their participation as a source of empowerment of their own preventive health care management and knowledge. In an RCT to decrease obesity risk among Black and Latino children (aged 9-12 years), DeFrank et al [16] reported assigning dedicated research staff as one of the key strategies. By doing so, DeFrank et al [16] noted that 81% of children at 3 months postintervention were very satisfied or extremely satisfied with how the study staff communicated and interacted with them.

Accommodation of Time, Place, and Transportation

Korean Americans who are first-generation immigrants experience busy immigration lifestyles, involving several family obligations, financial responsibilities, and social status, because of which health care management is often neglected [25]. Several church members refused to participate in the e-CHECC-uP trial because of their parental and familial duties. When proposed with alternative midweek and weekend options, some eligible women stated that it was simply too inconvenient to schedule meetings because of their children's schedules. Likewise, 4 out of 6 participants who withdrew from the study shared that their main reason was time constraints or that the study was *too burdensome* to continue. A participant stated that "it was a bit overwhelming...to make time for the phone counseling since I have other work to do," which further shows the participant's lack of time because of a busy lifestyle, not because of their lack of interest in the study itself. To reach and better retain participants, our research team accommodated controllable measures such as study visits, call times, and meeting places according to the individual preferences of the study participants. For example, most study visits for both recruitment and follow-up data collection were made on Sundays at a church as a preferred and most convenient community location for the participants. DeFrank et al [16] also noted that offering flexible hours was one of the factors associated with a high retention rate of 84%. Flexible scheduling was further enhanced by a web-based survey (Qualtrics) for data collection. We decided to use Qualtrics to allow willing participants to complete their follow-up study surveys remotely, at their own time and convenience. Qualtrics proved to be a viable resource, especially with the onset of COVID-19. In March 2020, we halted all in-person research visits to prioritize safety, given the increased risk of exposure to COVID-19. Despite having the added responsibility of uploading voice recordings for the reading test section for health literacy assessments, participants in the e-CHECC-uP trial were compliant in collaborating via Qualtrics. Using Qualtrics, we were able to retain all remaining participants at the onset of COVID-19 and achieved a retention rate of 85% (34/40).

Financial Incentives

We used an electronic gift card to partially compensate participants for their time spent on the study activities. A participant noted her experience:

> *If I participate, I get something in return. It is a total motivation.*

Another participant shared that it was a good side income to buy lunch for her children. In a systematic review of 56 studies focusing on African American participation in eHealth or mHealth intervention studies, James et al [21] found that 25 of the included studies offered some type of incentive, ranging from monetary (US $10-US $175) to nonmonetary incentives such as prepaid cell phones, free hypertension medication, and gym membership. Although the studies that provided incentives yielded a higher retention rate, James et al [21] recommended that researchers carefully determine the type and structure of incentives tailored to specific populations, communities, and targeted health behaviors. DeFrank et al [16] also noted that incremental monetary compensation through gift cards, which was selected based on formative research by surveying participants and the community, was an effective recruitment and retention strategy.

Diversity Among the Study Team

Language barriers remain one of the prevailing challenges among immigrants and are often associated with socioeconomic disadvantage, higher poverty, and uninsured rates. In 2012, 34%

XSL•FO
RenderX

of the immigrant population in the United States was limited English proficient compared with 9% of the total population. According to the Association of Asian Pacific Community Health Organizations [26], Korean Americans reported the second highest (46%) limited English proficient rates, followed by Vietnamese (53%), and had the highest rate of reporting poor patient-physician communication (34%) compared with other subethnic groups. Our findings were similar, with 50% (20/40) of our participants reporting that they were unable to make phone calls in English and 58% (23/40) reporting that they would not be able to go to an American hospital without an interpreter. Our study team members were diverse in terms of Korean fluency, technological competency, and gender. In particular, the bilingual study staff were instrumental in developing partnerships with study sites and community resources, disseminating critical information such as details on the consent forms during recruitment, and providing phone counseling for participants. A technology-based study on the practical issues in recruitment and retention of racial and ethnic minorities reported that many of their study participants were only comfortable participating in the study because they could use their heritage languages and were coached or supported in their heritage languages throughout the intervention [18]. In addition, the technologically inclined staff on the e-CHECC-uP study team were vital for troubleshooting issues with the web-based education modules, developing user-friendly remote surveys with prerecorded web-based instructions, and facilitating lines of e-communication with research participants via automated text messaging. Although 90% (36/40) of our participants reported having access to the internet at baseline, 70% (28/40) of the intervention women agreed or strongly agreed that they would need the support of a technical person to be able to use the product. Finally, despite initially operating with the notion that the study population would best respond to female study team members, as the e-CHECC-uP study focuses on women's cancer screening, we discovered that there were no observable differences in participant interaction or receptivity with male study team members. An internet-based study reported that racial and ethnic minority women responded well to race-concordant research assistants [20]. Therefore, we believe that trust was built between the study participants and the study team members, as they were of the same race and ethnicity.

Multiple, Yet Standardized, Modes of Communication

We used multiple modes of communication to accommodate individual communication preferences. Although phone calls were initially the main method of communication, we expanded our platforms to include text messaging and email after having failed to reach some study participants. A recent systematic review [19] revealed that adding email and telephone support to existing internet-based mental health interventions led to high completion rates and favorable mental health outcomes. Overall, text messaging was the most effective and preferred method to reach Korean American women in the e-CHECC-uP trial because it is quick to read and more accessible than phone calls or email. In fact, 39 out of 40 participants (97.5%) at baseline reported that they text at least once a week, and 34 participants (85%) reported that they text daily. Consequently, we scheduled general updates and reminders to be sent via text messaging using Qualtrics. We set up a study visit reminder protocol that involved an automated text message related to our scheduled follow-up data collection visit on Sunday that was sent out to participants a week before; participants would respond Y (yes) or N (no) to the text message to indicate their availability. If they responded Y, the study team would send a reminder text message on Friday and another on Sunday morning. Only when participants responded N were they contacted via phone to schedule a separate visit or to complete the study survey on the web. We received a 38% response rate using Qualtrics for 3-month and 6-month reminder text messages. DeFrank et al [16] also reported that a well-designed, standardized communication system was a key recruitment and retention strategy. The study staff used a single phone number for all calling and texting, a simplified study email address accessible to all study staff, and a single calendar as a central scheduling platform. All avenues of communication with the research participants were monitored and evaluated every week to develop improvement strategies as challenges surfaced. These interventions contributed to high survey completion rates throughout the study (84% at 3-months postintervention), high satisfaction rates (81%), and positive study experiences (97% of the participants felt that they received enough study information, and 80% (32/40) of the participants felt that their questions were answered properly).

Mobilizing Existing Community Relationships With Efforts to Build Trust

Before active recruitment in July 2019, the study team retrieved a list of ethnic churches in the Baltimore-Washington metropolitan area from the principal investigator's previous studies with the Korean American population. Additional churches from Google searches and word-of-mouth were identified. The word-of-mouth strategy is a well-known method for participant recruitment [27-30]. In addition, we mobilized a community health worker who had worked with the principal investigator in previous research studies. A community health worker is a "frontline public health worker who is a trusted member of and/or has an unusually close understanding of the community served" [31]. The involvement of community consultants and leaders in research not only increased positive recruitment results but also provided vital advice on how to approach participants and gain their trust [15]. Our community health worker helped us to build rapport with the pastor and church members, provided a brief introduction about the study, gathered those who showed interest, and advised on the timing of visits and how to approach potential participants. In addition, the active involvement of community leaders was impactful in recruiting and retaining participants for the e-CHECC-uP trial. For example, some pastors made verbal announcements at the end of the church service going above and beyond just posting our study information on their church bulletin board. Our study team members reflected and agreed that the participants from those churches were more engaging, asking questions, and being responsive to our contact. Our study team made additional efforts to build a collaborative relationship with the church. We were in consistent communication with the pastors via timely updates and greeting emails. This was especially effective in the beginning phase of the COVID-19 pandemic, as we

XSL•FO
RenderX

periodically checked in with our study participants and sent emails about our study progress and next steps in the face of uncertainty and restricted study site visits. To this end, we plan to disseminate our study findings to the churches, as we highlighted in our study completion letter to the participants. Sharing study results with community participants may further reinforce ethics and communication between the research and the community, reducing the stigma against clinical trials and population [32].

Prioritizing Features of Web-Based Intervention

Before the active recruitment of the e-CHECC-uP study participants, we conducted a usability test with 8 Korean American women who shared similar characteristics to the inclusion criteria used in the pilot clinical trial. For 5 months, we met with the women biweekly to hold their feedback to each developmental stage of the e-CHECC-uP educational website with the website developer. The features of web-based intervention programs were assessed to identify the ease and convenience of use, perceived helpfulness, and likelihood of future use. According to a recent systematic review, prioritizing the web-based intervention development enhanced participants' engagement [19]. The amount of time that participants committed to complete web-based activities was associated with the perceived study burden. In a study, a short average length of time lowered the participant burden and resulted in a high satisfaction rate [16]. Moreover, we provided a 1-sheet user guide to the participants in the intervention arm. This document was also available in Korean and English. As it summarizes the intervention module and highlights important features to increase participants' understanding and knowledge on the research topic, developing thorough written guidelines regarding interventions with technology is essential [15].

Combined Use of Web-Based and Direct Recruitment

Challenges with direct recruitment and enrollment, such as time constraints, motivated us to expand our recruitment approaches using a web-based recruitment strategy. Our web-based recruitment efforts yielded 100% success when they were used with interpersonal engagement. Interpersonal engagement in a community-based approach at the recruitment stage is vital for high enrollment and retention rates [21]. We contacted our web-based participants through previously established lines of in-person contact, engaged and communicated with them via texts and emails, and enrolled them through Qualtrics. This multilevel approach of using both web-based and direct recruitment allowed us to meet our target sample size and resulted in a more diverse participant pool, as this strategy was successful for relatively younger, bilingual participants who were in different life stages than the majority of our participants. Similarly, a systematic review [21] of mHealth intervention studies with African Americans suggested that a multilevel recruitment approach that considers culture, learning styles, and personal engagement may result in a larger, more diverse participant pool than using a single recruitment strategy.

Self-Directed Web-Based Intervention With Human Support

As more internet-based education and materials are offered, the importance of human support can be overlooked. However, human support and privacy or anonymity are still valued when delivering web-based interventions [19]. The e-CHECC-uP study included exclusive web-based intervention, including websites with videos on expected conversations and background information regarding mammograms and cervical cancer screening. Participants still found our study team members' assistance and monthly phone counseling to be helpful and vital to successful screening; 12 of 14 participants from the intervention group who completed the 6-month survey claimed that human support from the study team, phone counseling specifically, helped them to complete the screenings. A participant stated that the phone counseling reminded her of the importance of cancer screenings, "these studies are for my own benefits," and she preferred talking with the study team rather than a trip to the obstetrician-gynecologist to have "deep conversation," as her "secrets are being kept." Adding bilingual study team members with similar cultural backgrounds increases recruitment rates and has easier access to building trust and rapport with the participants; it also promotes acknowledging the importance of web-based interventions, as the study team can include culturally inclusive examples for better understanding [18]. Providing human support via phone counseling to remind the participants about web-based resources assisted in keeping participants in the loop but still maintaining social distancing to protect them.

Discussion

Principal Findings

We successfully recruited and retained Korean American women in a community-based cancer prevention pilot trial with a web-based intervention. The study team used diverse strategies to promote Korean American women's participation in the e-CHECC-uP trial. Our study's success was largely dependent on the collaboration between the study staff and the research participants. On the basis of our analysis of the e-CHECC-uP study and other relevant studies using web-based interventions, 9 categories of recruitment and retention strategies emerged. Although we did not measure the efficacy of each individual strategy, our thematic analysis resulted in particularly rich themes related to strategies addressing the quality of the bilingual research team.

Overall, we can attest that genuine care and accommodations are among the key strategies to engage participants and enhance their acceptance of our web-based study. Postintervention interviews supported the e-CHECC-uP study participants' trust in the study team and our genuine care. Mistrust has been noted as a major barrier to racial and ethnic minority participation in clinical trials [6]. Other successful approaches used to build trust include social networking through community engagement, such as working with community advisory boards, community health workers, and community leaders or gatekeepers [30,33-37]. In a community-based cervical cancer prevention study for African immigrant women, Cudjoe et al [36] provided

XSL•FO
RenderX

education about the research process and confidentiality to promote trust in research and research teams. These recruitment methods further extend to key retention strategies, such as developing trust-building relationships and realizing the benefits associated with an intervention among participants [33,38]. Further research is needed to investigate factors contributing to differences in perceptions of trust among community stakeholders, including whether expectations regarding long-term involvement impact trust.

Active communication with study participants resulted in a high retention rate (84%) and high participant satisfaction with the study [16]. Of the 75 children who completed the survey, 68% were very satisfied or satisfied with their experience in the study, and 81% of the children were extremely satisfied or satisfied with how the study staff communicated and interacted with them [16]. Similarly, we made deliberate efforts to constantly communicate with our participants and church leaders and to disseminate study progress and findings amid the COVID-19 pandemic, ultimately establishing trust and enduring relationships. Dissemination of study findings to bring the data back to the community is an important step in the research process, particularly for trials involving hard-to-reach communities. For example, according to a survey of community members (N=226) involved in health research with academic institutions, 73% of respondents noted disseminating research results to the community in a culturally relevant and appropriate manner as one of the most relevant success indicators of community engagement [39]. Future web-based trials involving racial and ethnic minorities should consider dissemination of study progress and findings as a means to engage study participants and other stakeholders in the research process, while promoting trust in the study team.

Web-based recruitment strategies using social media platforms such as WhatsApp and Facebook have been popular for recruiting African Americans, Hispanics, and Asians [22,36]. As COVID-19 is widespread and affects potential participants' lives, researchers should consider web-based methods to approach and retain participants. Without COVID-19, evidence suggests that the combined use of web-based recruitment with in-person strategies such as face-to-face contact and follow-up [33,35], word-of-mouth [27-30], and community-level recruitment [30,40,41] have been the most successful in recruiting racial and ethnic minorities. In addition to bilingual capacity and cultural knowledge, the e-CHECC-up team mobilized ethnic churches as the main setting for direct recruitment. Faith-based organizations serve as epicenters for social, religious, and health promotion activities across diverse racial and ethnic minority and immigrant communities, including African American, Hispanic, and Asian Americans [33,36,42,43]. Building upon existing community strengths and resources, faith-based organizations are well-suited community partners for health research and for long-term and sustainable collaboration [42]. Future web-based intervention trials should consider faith-based organizations as a key setting for the direct recruitment of racial and ethnic minorities.

Although several categories of recruitment and retention strategies identified from our analysis were also relevant to traditional in-person intervention programs, a few were unique to web-based interventions. Specifically, we found that prioritizing features of web-based intervention or use of human support were important to promote recruitment and retention of racial and ethnic minorities for web-based intervention trials. Some effective methods to create a relevant, useful web-based intervention are to work with target users during the development process [19]. Assessment of the usability of web-based intervention programs before launching the program may also increase the likelihood of its future use [44]. As more research trials progress into web-based interventions, it is important to integrate other facilitators such as cultural congruence, financial or medical benefits, convenience, and low risk in participation when proposing recruitment and retention strategies among racial and ethnic minorities [6].

Limitations

This study has several limitations. Our study used multiple recruitment and retention strategies. Consequently, we were unable to distinguish the most successful strategies in recruiting and retaining the target sample. For example, some of our participants reported that the financial incentives contributed to their motivation to join our study, and 1 participant stated that the cash incentives were used to buy basic necessities such as her child's school lunch. However, we did not have established metrics to measure the significance and impact of cash incentives on recruitment and retention efforts. Nevertheless, evidence strongly supports the use of multiple strategies to maximize recruitment and retention yields among racial and ethnic minorities [27-30,33,35,40,41]. The number of relevant web-based intervention trials included in our discussion might have been reduced because of our focus on recruitment and retention strategies. Similarly, the e-CHECC-uP study included a small and convenient sample of Korean American women. We expanded our discussion of recruitment and retention strategies by including other relevant studies using web-based interventions. Finally, we included studies conducted in the United States only. Therefore, the findings from the analysis may not be generalizable to racial or minority groups in other countries.

Conclusions

In conclusion, this analysis offers useful information regarding the recruitment and retention of racial and ethnic minority populations in web-based intervention trials. There is still a dearth of literature discussing whether and how racial and ethnic minorities have been recruited in clinical trials using technology-driven intervention approaches. This study contributes to the limited body of evidence that widely known strategies are essential and constructive in recruiting and retaining minority research participants. Future studies may include exploring successful recruitment and retention methods between race and ethnic or age and generation groups regarding their various internet penetration rates, comparing different recruitment strategies with or without using technology, and assessing their efficacy.

XSL•FO
RenderX

Acknowledgments

This study was supported by a grant from the Johns Hopkins Discovery Award Program. The authors would like to acknowledge Steve Chae, Veronica Lee, Jane Hong, and Lauren Tsao-Wu for their help with recruitment and retention efforts. Finally, the authors deeply appreciate all the study participants for their time and active participation in the trial. Without their volunteering, the study would not have been possible.

Conflicts of Interest

None declared.

References

1. McElfish PA, Long CR, Selig JP, Rowland B, Purvis RS, James L, et al. Health research participation, opportunity, and willingness among minority and rural communities of Arkansas. Clin Transl Sci 2018 Sep;11(5):487-497 [FREE Full text] [doi: 10.1111/cts.12561] [Medline: 29772113]
2. NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research. National Institute of Health. 2017. URL: https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm [accessed 2020-04-09]
3. Guerrero S, López-Cortés A, Indacochea A, García-Cárdenas JM, Zambrano AK, Cabrera-Andrade A, et al. Analysis of racial/ethnic representation in select basic and applied cancer research studies. Sci Rep 2018 Sep 18;8(1):13978 [FREE Full text] [doi: 10.1038/s41598-018-32264-x] [Medline: 30228363]
4. American Community Survey 1-year Estimates. United States Census Bureau. 2018. URL: https://data.census.gov/cedsci/table?q=race&d=ACS%201-Year%20Estimates%20Detailed%20Tables&tid=ACSDT1Y2018.B02001&hidePreview=true&y=2018&tp=true&moe=false&t=Race%20and%20Ethnicity [accessed 2020-04-10]
5. United States Population Projections: 2000 to 2050. United States Census Bureau. 2000. URL: https://www.census.gov/content/dam/Census/library/working-papers/2009/demo/us-pop-proj-2000-2050/analytical-document09.pdf [accessed 2020-04-09]
6. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]
7. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]
8. Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242 [FREE Full text] [doi: 10.1002/cncr.23157] [Medline: 18008363]
9. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010;15(Suppl 3):216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
10. Roser M, Ritchie H, Ortiz-Ospina E. Our World in Data. World Bank. URL: http://ourworldindata.org/internet [accessed 2020-04-09]
11. Internet/Broadband Fact Sheet. Pew Research Center. 2019. URL: https://pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2020-08-10]
12. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54 [FREE Full text] [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
13. Recommendations for the Early Detection of Breast Cancer. American Cancer Society. URL: https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer [accessed 2020-06-19]
14. Guidelines for the Prevention and Early Detection of Cervical Cancer. The American Cancer Society. 2018. URL: https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/cervical-cancer-screening-guidelines.html [accessed 2020-06-19]
15. Chee W, Ji X, Kim S, Park S, Zhang J, Chee E, et al. Recruitment and retention of asian americans in web-based physical activity promotion programs: a discussion paper. Comput Inform Nurs 2019 Sep;37(9):455-462 [FREE Full text] [doi: 10.1097/CIN.0000000000000541] [Medline: 31518339]
16. DeFrank G, Singh S, Mateo KF, Harrison L, Rosenthal A, Gorman A, et al. Key recruitment and retention strategies for a pilot web-based intervention to decrease obesity risk among minority youth. Pilot Feasibility Stud 2019;5:109 [FREE Full text] [doi: 10.1186/s40814-019-0492-8] [Medline: 31516726]
17. Staffileno BA, Zschunke J, Weber M, Gross LE, Fogg L, Tangney CC. The feasibility of using facebook, craigslist, and other online strategies to recruit young African American women for a web-based healthy lifestyle behavior change intervention. J Cardiovasc Nurs 2017;32(4):365-371. [doi: 10.1097/JCN.0000000000000360] [Medline: 27428356]

XSL•FO
RenderX

18. Im E, Kim S, Xu S, Lee C, Hamajima Y, Inohara A, et al. Issues in recruiting and retaining Asian American breast cancer survivors in a technology-based intervention study. Cancer Nurs 2020;43(1):E22-E29 [FREE Full text] [doi: 10.1097/NCC.0000000000000657] [Medline: 30346330]
19. Reyes AT, Constantino RE, Arenas RA, Bombard JN, Acupan AR. Exploring challenges in conducting e-mental health research among Asian American women. Asian Pac Isl Nurs J 2018;3(4):139-153 [FREE Full text] [doi: 10.31372/20180304.1016] [Medline: 31037262]
20. Im E, Lee Y, Ji X, Zhang J, Kim S, Chee E, et al. Internet recruitment of Asian American breast cancer survivors. ANS Adv Nurs Sci 2016;39(3):E17-E27 [FREE Full text] [doi: 10.1097/ANS.0000000000000131] [Medline: 27490884]
21. James DC, Harville C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-health and m-health studies: a systematic review. Telemed J E Health 2017 May;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]
22. Tsai W, Zavala D, Gomez S. Using the Facebook advertisement platform to recruit Chinese, Korean, and Latinx cancer survivors for psychosocial research: web-based survey study. J Med Internet Res 2019 Jan 10;21(1):e11571 [FREE Full text] [doi: 10.2196/11571] [Medline: 30632966]
23. Ethical, legal, and regulatory framework for human subjects research. In: Optimizing the Nation's Investment in Academic Research: a New Regulatory Framework for the 21st Century. Washington, DC: National Academies Press; 2016.
24. Woolf SH, Zimmerman E, Haley A, Krist AH. Authentic engagement of patients and communities can transform research, practice, and policy. Health Aff (Millwood) 2016 Apr;35(4):590-594 [FREE Full text] [doi: 10.1377/hlthaff.2015.1512] [Medline: 27044956]
25. Nam S, Song H, Park S, Song Y. Challenges of diabetes management in immigrant Korean Americans. Diabetes Educ 2013;39(2):213-221. [doi: 10.1177/0145721713475846] [Medline: 23427240]
26. Association of Asian Pacific Community Health Organizations. 2014. URL: https://www.aapcho.org/wp/wp-content/uploads/2014/08/AANHOPI-LEP-Fact-Sheet_2014_final.pdf [accessed 2020-06-18]
27. Maxwell AE, Bastani R, Vida P, Warda US. Strategies to recruit and retain older Filipino-American immigrants for a cancer screening study. J Community Health 2005 Jun;30(3):167-179 [FREE Full text] [doi: 10.1007/s10900-004-1956-0] [Medline: 15847243]
28. Wang JH, Sheppard VB, Liang W, Ma GX, Maxwell AE. Recruiting Chinese Americans into cancer screening intervention trials: strategies and outcomes. Clin Trials 2014 Apr;11(2):167-177 [FREE Full text] [doi: 10.1177/1740774513518849] [Medline: 24567288]
29. Hanza MM, Goodson M, Osman A, Porraz Capetillo MD, Hared A, Nigon JA, et al. Lessons learned from community-led recruitment of immigrants and refugee participants for a randomized, community-based participatory research study. J Immigr Minor Health 2016 Oct;18(5):1241-1245 [FREE Full text] [doi: 10.1007/s10903-016-0394-2] [Medline: 26984117]
30. Lopez-Owens M, Starkey K, Gil C, Armenta K, Maupomé G. The vidasana study: recruitment strategies for longitudinal assessment of egocentric hispanic immigrant networks. Int J Environ Res Public Health 2018 Dec 15;15(12):- [FREE Full text] [doi: 10.3390/ijerph15122878] [Medline: 30558280]
31. Community Health Workers. American Public Health Association. URL: https://www.apha.org/apha-communities/member-sections/community-health-workers [accessed 2020-07-18]
32. Chen PG, Diaz N, Lucas G, Rosenthal MS. Dissemination of results in community-based participatory research. Am J Prev Med 2010 Oct;39(4):372-378. [doi: 10.1016/j.amepre.2010.05.021] [Medline: 20837290]
33. Choi E, Heo GJ, Song Y, Han H. Community health worker perspectives on recruitment and retention of recent immigrant women in a randomized clinical trial. Fam Community Health 2016;39(1):53-61 [FREE Full text] [doi: 10.1097/FCH.0000000000000089] [Medline: 26605955]
34. Luebbert R, Perez A. Barriers to clinical research participation among African Americans. J Transcult Nurs 2016 Sep;27(5):456-463. [doi: 10.1177/1043659615575578] [Medline: 25754929]
35. Mendez-Luck CA, Trejo L, Miranda J, Jimenez E, Quiter ES, Mangione CM. Recruitment strategies and costs associated with community-based research in a Mexican-origin population. Gerontologist 2011 Jun;51 Suppl 1:S94-105 [FREE Full text] [doi: 10.1093/geront/gnq076] [Medline: 21565824]
36. Cudjoe J, Turkson-Ocran R, Ezeigwe AK, Commodore-Mensah Y, Nkimbeng M, Han H. Recruiting African immigrant women for community-based cancer prevention studies: lessons learned from the Afropap study. J Community Health 2019 Oct;44(5):1019-1026 [FREE Full text] [doi: 10.1007/s10900-019-00677-y] [Medline: 31102117]
37. Lim HJ, Lin C, Liu Y, Chee W, Im E. Effective strategies for recruiting of Asian cancer patients in internet research. Stud Health Technol Inform 2006;122:882. [Medline: 17102441]
38. Sankaré IC, Bross R, Brown AF, Del Pino HE, Jones LF, Morris DM, et al. Strategies to build trust and recruit African American and Latino community residents for health research: a cohort study. Clin Transl Sci 2015 Oct;8(5):412-420 [FREE Full text] [doi: 10.1111/cts.12273] [Medline: 26094679]
39. Skinner JS, Williams NA, Richmond A, Brown J, Strelnick AH, Calhoun K, et al. Community experiences and perceptions of clinical and translational research and researchers. Prog Community Health Partnersh 2018;12(3):263-271 [FREE Full text] [doi: 10.1353/cpr.2018.0050] [Medline: 30581169]

XSL•FO
RenderX

40. Wong CK, Horn-Ross PL, Gee GC, Shariff-Marco S, Quach T, Allen L, et al. Strategies for recruiting representative samples of Asian Americans for a population-based case-control study. J Epidemiol Community Health 2016 Oct;70(10):974-982. [doi: 10.1136/jech-2015-206905] [Medline: 27053683]
41. Coronado GD, Ondelacy S, Schwarz Y, Duggan C, Lampe JW, Neuhouser ML. Recruiting underrepresented groups into the Carbohydrates and Related Biomarkers (CARB) cancer prevention feeding study. Contemp Clin Trials 2012 Jul;33(4):641-646 [FREE Full text] [doi: 10.1016/j.cct.2012.03.017] [Medline: 22504222]
42. Oppenheim CE, Axelrod K, Menyongai J, Chukwuezi B, Tam A, Henderson DC, et al. The HEAAL project: applying community-based participatory research (CBPR) methodology in a health and mental health needs assessment with an African immigrant and refugee faith community in Lowell, Massachusetts. J Public Health Manag Pract 2019;25(1):E1-E6 [FREE Full text] [doi: 10.1097/PHH.0000000000000707] [Medline: 30507809]
43. Metayer N, Boulos R, Tovar A, Gervis J, Abreu J, Hval E, et al. Recruitment of new immigrants into a randomized controlled prevention trial: the live well experience. J Prim Prev 2018 Oct;39(5):453-468. [doi: 10.1007/s10935-018-0519-6] [Medline: 30128810]
44. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]

Abbreviations

e-CHECC-uP: e-Community–Based Health Literacy–Focused Intervention for Cancer Control
mHealth: mobile health
RCT: randomized controlled trial

Edited by R Kukafka; submitted 31.08.20; peer-reviewed by A Reyes, EO Im, J Yang; comments to author 28.09.20; revised version received 26.10.20; accepted 30.04.21; published 12.07.21.

Please cite as:
Hwang DA, Lee A, Song JM, Han HR
Recruitment and Retention Strategies Among Racial and Ethnic Minorities in Web-Based Intervention Trials: Retrospective Qualitative Analysis
J Med Internet Res 2021;23(7):e23959
URL: https://www.jmir.org/2021/7/e23959
doi:10.2196/23959
PMID:34255658

©DaSol Amy Hwang, Alex Lee, Jae Min Song, Hae-Ra Han. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measuring Success of Patients' Continuous Use of Mobile Health Services for Self-management of Chronic Conditions: Model Development and Validation

Ting Song[1,2], MSc; Ning Deng[3], PhD; Tingru Cui[4], PhD; Siyu Qian[1,2,5], PhD; Fang Liu[6], MD; Yingping Guan[6], BSc; Ping Yu[1,2], PhD

[1]Centre for Digital Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia

[3]The Ministry of Education Key Laboratory of Biomedical Engineering, College of Biomedical Engineering and Instrument Science, Zhejiang University, Huangzhou, China

[4]School of Computing and Information Systems, Faculty of Engineering and Information Technology, University of Melbourne, Melbourne, Australia

[5]Drug and Alcohol Service, Illawarra Shoalhaven Local Health District, Wollongong, Australia

[6]Department of Health Examination, General Hospital of Ningxia Medical University, Yinchuan, China

Corresponding Author:
Ping Yu, PhD
Centre for Digital Transformation, School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 2 4221 5412
Fax: 61 2 4221 4170
Email: ping@uow.edu.au

Abstract

Background: Mobile health services are gradually being introduced to support patients' self-management of chronic conditions. The success of these services is contingent upon patients' continuous use of them.

Objective: This study aims to develop a model to measure the success of patients' continuous use of mobile health services for the self-management of chronic conditions.

Methods: The proposed model was derived from the information systems continuance model and the information systems success model. This model contains 7 theoretical constructs: information quality, system quality, service quality, perceived usefulness, user satisfaction, perceived health status, and continuous use intention. A web-based questionnaire survey instrument was developed to test the model. The survey was conducted to collect data from 129 patients who used a mobile health app for hypertension management from 2017 to 2019. The questionnaire items were derived from validated instruments and were measured using a 5-point Likert scale. The partial least squares modelling method was used to test the theoretical model.

Results: The model accounted for 58.5% of the variance in perceived usefulness (R^2=0.585), 52.3% of the variance in user satisfaction (R^2=0.523), and 41.4% of the variance in patients' intention to make continuous use of mobile health services (R^2=0.414). The continuous use intention was significantly influenced by their perceived health status (β=.195, *P*=.03), perceived usefulness (β=.307, *P*=.004), and user satisfaction (β=.254, *P*=.04) with the mobile health service. Information quality (β=.235, *P*=.005), system quality (β=.192, *P*=.02), and service quality (β=.494, *P*<.001) had a significantly positive influence on perceived usefulness but not on user satisfaction. Perceived usefulness had a significantly positive influence on user satisfaction (β=.664, *P*<.001). In a result opposite to the original hypothesis, perceived health status did not negatively influence patients' intention to continue using the mobile health service but showed a significantly positive correlation.

XSL•FO
RenderX

Conclusions: This study developed a theoretical model to predict and explain patients' continuous use of mobile health services for self-management of chronic conditions and empirically tested the model. Perceived usefulness, user satisfaction, and health status contributed to patients' intention to make continuous use of mobile health services for self-managing their chronic conditions.

(J Med Internet Res 2021;23(7):e26670) doi:10.2196/26670

KEYWORDS

mobile health; service; smartphone; mobile application; continuous use; high blood pressure; chronic disease; PLS

Introduction

Background

Mobile health (mHealth) services have been increasingly introduced to support patients in their self-management of chronic conditions over the last decade [1,2]. mHealth overcomes the traditional barriers of time, distance, and cost by providing patients with access to health information, assessment, and assistance anytime and anywhere [3,4]. As one of the most used mHealth services, mHealth apps are widely used to record and evaluate patients' vital signs and self-management behaviors such as medication, exercise, and diet; access health information; remind self-management behaviors; and communicate with health care providers [1,2,5]. With increasing awareness and ability, patients are motivated to self-manage their behavior, that is, adhere to treatment, thereby controlling their chronic conditions and maintaining the quality of life. When used over a long period of time, these apps are likely to result in positive outcomes such as obvious changes in health behaviors [6-8]. Despite their potential benefits, mHealth apps are rarely used. Perez [9] found that 25% of mHealth apps were used only once after installation, and most users stopped using these apps after 4 periods of interaction with the apps. A national survey conducted in the United States showed that 45.7% (427/934) of the participants who downloaded certain mHealth apps no longer used these apps [10]. This is contrary to the original intention of introducing these apps to help the management of long-term diseases because short-term use is not sufficient to achieve the expected benefits [11,12]. Therefore, it is essential to understand the factors that affect patients' continuous use of mHealth apps.

Prior studies on patients' use of mHealth apps have focused on patient acceptance and initial use of these apps [13-22], whereas only a few studies have focused on the continuous use of mHealth services [23-26]. Some of these continuous use studies only examined mHealth services for general health care, such as appointments and health consultations instead of self-management of chronic conditions [23,24,26]. Cho [26] developed and tested a model to explain the mechanism that determines the continuous use intention of mHealth services. He found that continuous use intention is influenced by confirmation of the primary expectation of mHealth apps but he did not specify the content of the expectation. Lee et al [25] found that patients' intention to continuously use an mHealth service was closely associated with their regular use of self-monitoring functions; however, they did not further explore the relationship between these 2 constructs.

To date, there has been little theoretical research on the factors influencing patients' continuous use of mHealth apps to support self-management of chronic conditions. To fill this gap, this research aims to (1) identify the influencing factors and develop a predictive model to explain their relationships with patients' continuous use of mHealth apps for self-management of chronic conditions, (2) develop and validate a questionnaire survey instrument that empirically tests and theorizes the model, and (3) examine the associations among the variables and their relative impact on the continuous use of an mHealth app. The theoretical model was tested in the context of hypertension self-management.

Theoretical Foundation

The constructs of the proposed model are drawn from the information systems (IS) continuance model [27] and IS success model [28]. Inspired by Oliver's Expectation Confirmation Theory [29], Bhattacherjee [27] proposed that information system users' continuous use intention is similar to their repurchase decision-making. They compared the benefits acquired from using an information system product or service with the prior expectation to decide the level of satisfaction with it. The comparison result, that is, confirmation, became their reference for continuous use. This theory is known as the IS continuance model. In this model, perceived usefulness and user satisfaction are considered crucial factors influencing continuous use intention and are associated with confirmation, thus being deemed intermediate variables. However, the content of the confirmation is not clearly defined in this model, which makes it difficult to assess the factors that determine the continuous use intention.

In their famous IS success model, Delone and McLean [28] proposed that 3 factors, namely, information quality, system quality, and service quality determine use intention and satisfaction, and use intention and satisfaction predict information system success. This theory has been widely adopted in studies that evaluate information system success [30,31]. According to a published mobile technology acceptance model, usefulness predicts use intention [13]. Therefore, we posit that information quality, system quality, and service quality may affect perceived usefulness and user satisfaction of mHealth services for patient self-management of chronic conditions.

Research Hypotheses

Information Quality, System Quality, Service Quality, and Perceived Usefulness

Information quality refers to the quality of content that the mobile service provides. Its attributes include relevance, timeliness of update, and ease of understanding [28,32]. Perceived usefulness refers to users' ex-post expectations and beliefs about the effectiveness and benefits of using an mHealth

XSL•FO
RenderX

app from their experience [27]. Information quality is one of the critical determinants for information system success because the acquisition of information is the main purpose for users to use an information system. Alsabawy et al [33] found that low-quality information provided by an e-learning system may mislead users and consequently change their perception of its usefulness. System quality refers to the overall performance of an mHealth app as perceived by the users [28,32]. It measures the technical success of an mHealth service. As an information system is the carrier of information, its quality is the prerequisite for ensuring that users can easily obtain the information they need. For example, if the functions of an information system are too complex and difficult to use, users may not invest time and energy to learn and use the system, which may weaken users' perception of the usefulness of the system [30]. Service quality refers to the support that an mHealth app user can receive from the support personnel and technical team who administer the system [28,32]. The attributes include dependability, availability, and empathy of the support staff. Dou et al [13] found that a good doctor-patient relationship was one of the prerequisites for patients to think that an mHealth service was useful to them. For example, they thought the mHealth service was useful only if the patients knew the health care provider could help them solve problems through the app. Wu [31] also found that high-quality service can increase users' perceived usefulness of a web-based health care community. Watts et al [34] suggested that assessing the relationships between these information system factors needs to consider the context of information system use. Therefore, the following hypotheses were posited:

Hypothesis 1(a): *Information quality is positively associated with patients' perceived usefulness of mHealth services.*

Hypothesis 1(b): *System quality is positively associated with patients' perceived usefulness of mHealth services.*

Hypothesis 1(c): *Service quality is positively associated with patients' perceived usefulness of mHealth services.*

Information Quality, System Quality, Service Quality, and User Satisfaction

User satisfaction refers to a user's emotional or psychological state about using a system [28]. Delone and McLean [28] posited that system quality, information quality, and service quality predict user satisfaction. For example, patients are dissatisfied with irrelevant information in the app, such as overwhelming advertisements [24]. Users prefer a clean and simple interface and easy-to-understand navigation menu. Zheng et al [32] reported that if the service provided by an information system does not provide the expected reliability or consistency, satisfaction with the information system will decrease. Consequently, the following hypotheses were posited:

Hypothesis 2(a): *Information quality is positively associated with user satisfaction of mHealth services.*

Hypothesis 2(b): *System quality is positively associated with user satisfaction of mHealth services.*

Hypothesis 2(c): *Service quality is positively associated with user satisfaction of mHealth services.*

Perceived Usefulness, User Satisfaction, Perceived Health Status, and Continuous Use Intention

Bhattacherjee [27] suggested that perceived usefulness predicts user satisfaction due to a likely positive emotional response derived from improvement in work efficiency and job performance by using the information system [35]. With this understanding, Kim and Lee [36] focused on investigating the usefulness of their robot services to improve user satisfaction. Likewise, if patients believe that using a mHealth app can help them control hypertension, they should also be satisfied with the app. Therefore, the following hypothesis was posited:

Hypothesis 3: *Perceived usefulness of mHealth services is positively associated with user satisfaction with such services.*

Perceived health status refers to individuals' assessments of their health conditions [37]. It reflects the physiological, behavioral, social, and psychological conditions that a person experiences, which can be difficult to capture by other objective indicators [38]. Song et al [8] found that many patients reduced usage frequency or even stopped using the mHealth service once their blood pressure was under control. However, they would resume use once their blood pressure increased again.

Bhattacherjee [27] showed that perceived usefulness and user satisfaction will positively predict continuous use intention. Vaghefi and Tulu [24] interviewed 17 people who used different apps to support their lifestyle changes, that is, promoting physical activity and mindfulness. The comparison of user perceptions 2 weeks before and after app usage indicated that user experience and perceived health goals are 2 factors influencing these people's continuous use intention. Therefore, the following hypotheses were posited:

Hypothesis 4(a): *Perceived health status is negativelyassociated with continuous use intention of mHealth services.*

Hypothesis 4(b): *Perceived usefulness is positively associated with continuous use intention of mHealth services.*

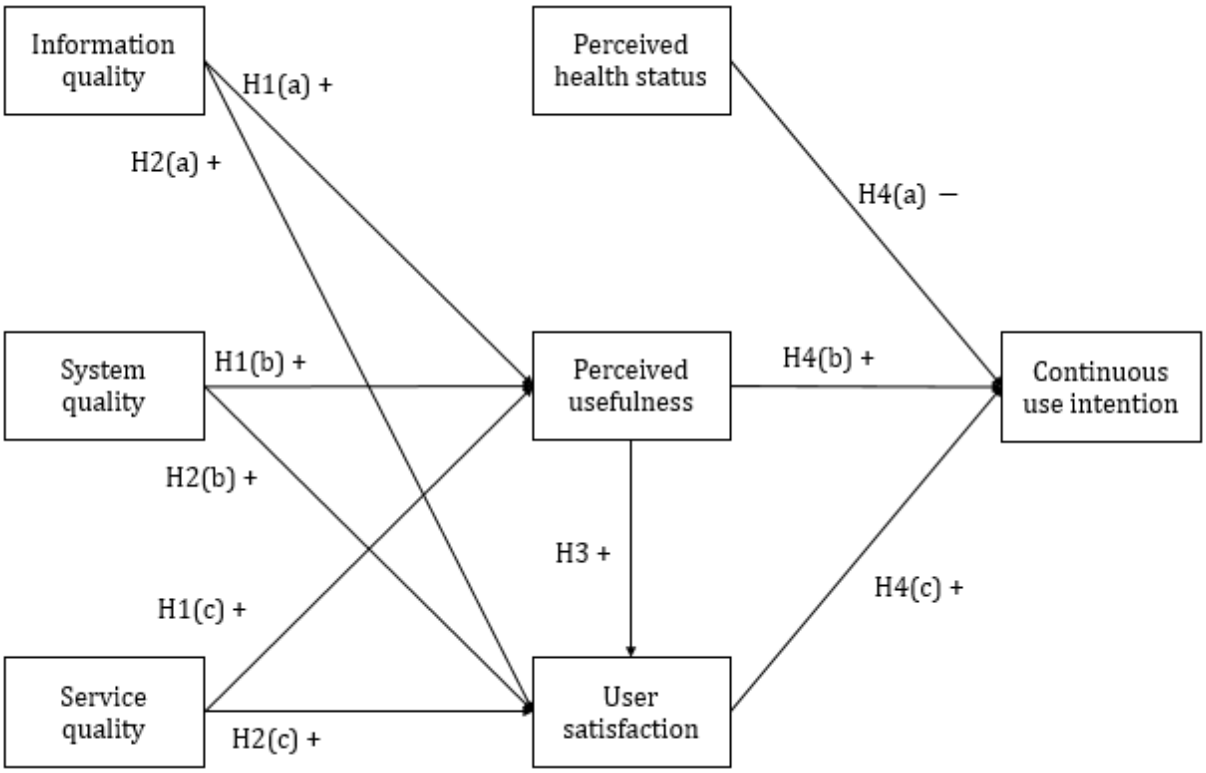
Hypothesis 4(c): *User satisfaction is positively associated with continuous use intention of mHealth services.*

The Proposed Theoretical Model

The proposed continuous use model of mHealth services for self-management of chronic conditions consists of 7 variables (Figure 1). Four of them are independent variables: information quality, system quality, service quality, and perceived health status. Continuous use intention is the dependent variable. The remaining 2 are mediators: perceived usefulness and user satisfaction.

XSL•FO
RenderX

Figure 1. The proposed research model. "H" refers to the hypotheses in this study. +: positive association; –: negative association.



Methods

Ethics Approval

This study was approved by the Human Research Ethics Committee of the General Hospital of Ningxia Medical University, China (Registration number: 2018-325).

Study Setting and mHealth Service

To improve patient hypertension management and population health, an international tripartite collaborative research program was piloted. The mHealth service entitled “Blood Pressure Assistant” was developed by the Biomedical Informatics Laboratory at Zhejiang University, China, to assist outpatients in self-managing hypertension. The mHealth service included an app for the patients to use and a web-based portal for health care providers, including clinicians and certified health managers, to communicate with the patients. The functional modules of the app were (1) digital forms to record vital signs, that is, blood pressure, heart rate, weight, and self-management behaviors, that is, medication, exercise, and diet; (2) health education materials; (3) reminders for self-management; (4) reports on daily/monthly statistical trend of vital signs and self-management performance; and (5) feedback about the blood pressure level, being normal or not. The functional modules of the web-based portal included (1) statistical results and visualization charts of data recorded by patients; (2) alerts on abnormal situations; and (3) patient classification and follow-up tracking reminders. The mHealth service was implemented in the Department of Cardiology, General Hospital of Ningxia Medical University, China, in November 2015. The effectiveness of this service was evaluated in a clinical trial (Registration number: ChiCTR1900026437). The trial participants were outpatients with hypertension of the hospital and were provided with a 1-hour face-to-face training before using the app. The training included the significance and methods of hypertension self-management, the way to install and use the app, and tips to solve common usage problems.

Questionnaire Survey Development and Data Collection

A self-administered questionnaire was developed in 2017 to collect data to measure the 7 latent variables by using a 5-point Likert scale (Table 1) and to test the relationships among them. The questionnaire consisted of 17 questions, and all were adopted from previously validated studies and modified to fit our study context. Although each latent construct is best measured by at least 3 items [39], in balancing rigor with avoiding survey fatigue for satisfactory response rates [40], we only used 2 items to measure 5 constructs in reference to the recommendation of previous literature [13,26,30]. Each question was anchored between 1 (strongly disagree) and 5 (strongly agree). The measurement items were translated into Chinese by one researcher and then discussed and validated by a panel of 9 bilingual experts, that is, 4 clinicians, 3 medical informatics experts, 1 certified health manager, and 1 information system expert. One researcher then back translated the instrument into English to confirm the accuracy and quality of the Chinese translation.

XSL•FO
RenderX

Table 1. Questionnaire constructs and their measurement items scored using a 5-point Likert scale.

Construct, item code	Item description [citations of validated studies]	Score, mean (SD)[a]
Information quality (IQ)		
IQ1	Information provided by the BP[b] Assistant is relevant for my needs [28,29,32].	4.310 (0.745)
IQ2	Information provided by the BP Assistant is sufficient for my needs [28,32].	4.318 (0.757)
IQ3	Information provided by the BP Assistant has been updated in a timely manner [28,32].	3.953 (0.879)
System quality (SysQ)		
SysQ1	The BP Assistant is easy to learn [29].	4.519 (0.572)
SysQ2	The BP Assistant is easy to use [28,29].	4.481 (0.624)
Service quality (SerQ)		
SerQ1	The health manager can resolve problems that I encountered when using the BP Assistant [28].	4.271 (0.745)
SerQ2	The health manager gives me individual guidance in the follow-up phone calls [28,32].	4.085 (0.778)
Perceived usefulness (PU)		
PU1	The BP Assistant is of benefit to me [26].	4.202 (0.730)
PU2	Overall, using the BP Assistant is advantageous to my hypertension management [27].	4.380 (0.587)
User satisfaction (US)		
US1	I am content with my use of BP Assistant [27,41,42].	4.504 (0.599)
US2	I am satisfied with my use of BP Assistant [27,29,41,42].	4.488 (0.544)
Perceived health status (PHS)		
PHS1	Currently, my BP is back to normal [8,43].	4.101 (0.796)
PHS2	Currently, my BP is under control [8,43,44].	4.031 (0.787)
PHS3	Currently, I don’t worry about my health [13].	4.225 (0.707)
Continuous use intention (CUI)		
CUI1	I intend to continue using the BP Assistant rather than discontinue its use [27,42].	4.519 (0.558)
CUI2	If I could, I would like to discontinue my use of the BP Assistant (reverse coded) [27].	4.116 (0.850)

[a]Scored on a 5-point Likert scale. Each question was anchored between 1 (strongly disagree) and 5 (strongly agree).

[b]BP: blood pressure.

The questionnaire survey was then built into the mHealth app as a special functional module for the trial participants to complete. A participant information sheet was placed at the beginning of the questionnaire to inform them about the purpose of the survey, the voluntary nature of completing it, and the confidentiality and anonymity of their responses for any derived research publication. The participants could click a checkbox to express their consent. Some participants returned the completed questionnaire without clicking this checkbox, and this was treated as implied consent. The questionnaire responses were extracted from the app database in March 2020 for analysis. The patients’ demographic information, that is, age, gender, and education level, was extracted from the database of the web-based portal.

Data Analysis

The research model was tested by partial least squares structural equation modelling (PLS-SEM) using the software program SmartPLS (version 3.0, SmartPLS GmbH) [39,45,46]. The PLS-SEM is commonly used to model the dynamic relationships between antecedent variables and dependent variables, thereby addressing the limitation of the multiple regression model with a relatively fixed relationship between variables and multicollinearity issues [47]. Moreover, the average number of latent variables in PLS-SEM is 7.94, which is much higher than 4.70 in covariance-based SEM [46]. Therefore, PLS-SEM is more conducive to solving more complex models. In addition, compared with the covariance-based SEM, PLS-SEM has a high level of statistical power even when the sample size is relatively small [48]. This is very practical for studies that have difficulties in recruiting research subjects, especially for those that are inherently complex and sensitive, like this study. The measurement model was tested by assessing reliability and validity [39,49]. For reflective constructs, the indicator reliability was assessed by the indicator loadings and collinearity statistics, that is, outer variance inflation factor (VIF) values. The construct reliability was assessed by Cronbach alpha and composite reliability. The convergent validity was assessed by the average variance extracted. The discriminant validity was assessed by the Fornell-Larcker criterion, cross-loading, and heterotrait-monotrait ratio [39,50]. For formative constructs, reliability was an irrelevant assessment criterion [49]. The

XSL•FO
RenderX

indicator validity was assessed by the indicator weights and VIF. The discriminate validity was assessed by interconstruct correlations. The structural model was tested by path coefficients (β), variance explained (R^2), effect size (f^2), and the blindfolding-based cross-validated redundancy measure (Q^2). Path coefficients (β) measured the direct effect of a variable assumed to be a cause on another variable assumed to be an effect [51]—a positive β value referred to a positive association and vice versa. Variance explained (R^2) referred to in-sample predictive power, which measured a model's explanatory power [52,53]. Effect size (f^2) explained the changes before and after an exogenous construct is included and excluded from the model [54]. The f^2 values of 0.02, 0.15, and 0.35 were considered as small, medium, and large effects, respectively. The Q^2 value was used to assess the PLS path model's predictive accuracy [55,56].

Results

Survey Participants' Characteristics

A total of 141 participants completed the questionnaire survey. Four of them gave the same answer to all the questions. Another 8 provided the same answer to questions 16 and 17, which were opposite to each other in nature. These responses were considered invalid responses and were excluded from further data analysis. Therefore, 129 responses were used in the statistical analysis. There are 2 requirements regarding the sample size. First, the ratio of the sample size to the number of parameters should be greater than 5:1 [57]; second, the sample size should be greater than 10 times the largest number of either the formative items used to measure a single construct or the largest number of paths the latent variable has in the model [45,47]. In this study, the number of parameters is 7, the number of formative items is 3, and the number of paths the latent variable has is 10. Therefore, our sample size of 129 responses surpassed the 2 threshold requirements. The participants' ages ranged from 34 to 79 years, with a median of 53 years. Two times more males than females participated in the survey. Approximately 69.8% (90/129) of the respondents were in the age group of 40-59 years, 62.8% (81/129) were in the workforce, and 75.9% (98/129) had an education level of high school or above (Table 2).

Table 2. Demographics of the participants (N=129).

Characteristics	Values, n (%)
Gender	
Male	98 (76.0)
Female	31 (24.0)
Age (years)	
<40	7 (5.4)
40-49	48 (37.2)
50-59	42 (32.6)
60-69	23 (17.8)
≥70	9 (7.0)
Employment status	
Employed	66 (51.2)
Self-employed	15 (11.6)
Unemployed	7 (5.4)
Retired	29 (22.5)
No response	12 (9.3)
Education level	
Primary and middle school	22 (17.1)
High school	25 (19.4)
University/college/graduate	69 (53.5)
Postgraduate	4 (3.1)
No response	9 (6.9)

XSL•FO
RenderX

Descriptive Statistics of the Constructs

The mean scores of all the latent variables were positive, that is, close to or over 4 in the 5-point Likert scale (Table 1). In particular, the first item to measure the continuous use intention, "I intend to continue using the Blood Pressure Assistant rather than discontinue its use" reached a high mean score of 4.519, suggesting that the participants had high intention to continue to use the mHealth service.

Measurement Model Validation

For the reflective constructs, that is, perceived usefulness, user satisfaction, perceived health status, and continuous use intention, each item was loaded above the threshold value of 0.708 on its respective construct and was significant at P=.01 (Table 3 and Multimedia Appendix 1). All VIF values were less than 3, indicating some correlation but not enough to be overly concerned about [58]. These confirmed the indicator reliability. All Cronbach alpha and composite reliability values were above .700, which confirmed the construct reliability. All average variance extracted values were more than 0.500, with their square root presented on the diagonal (Table 4). The average variance extracted of each construct was greater than its squared correlations with other constructs, confirming the convergent validity. The cross-loadings of each indicator on other constructs were lower than that on its designated construct, and each indicator loaded highest on its own construct. All the heterotrait-monotrait values were below 0.900. These confirmed the discriminant validity of the constructs [59]. For the formative constructs, that is, information quality, system quality, and service quality, each item was weighted above the threshold value of 0.200 on its respective construct and was significant at P=.01. All the VIF values were below 3, confirming indicator validity [58]. The correlations between the formative and all the other constructs were less than 0.700, confirming discriminant validity.

Table 3. Descriptive statistics of the construct, internal reliability, and convergent validity.

Construct and scale, item code	Standardized loading/weight	Variance inflation factor (outer)	Cronbach alpha	Composite reliability	Average variance extracted
Information quality (IQ), Formative			N/A[a]	N/A	0
IQ1	0.292	1.632			
IQ2	0.555	1.513			
IQ3	0.395	1.284			
System quality (SysQ), Formative			N/A	N/A	0
SysQ1	0.428	2.032			
SysQ2	0.649	2.032			
Service quality (SerQ), Formative			N/A	N/A	0
SerQ1	0.369	2.284			
SerQ2	0.693	2.284			
Perceived usefulness (PU), Reflective			.829	0.921	0.854
PU1	0.917	2.006			
PU2	0.931	2.006			
User satisfaction (US), Reflective			.787	0.903	0.824
US1	0.900	1.725			
US2	0.915	1.725			
Perceived health status (PHS), Reflective			.771	0.867	0.684
PHS1	0.828	1.759			
PHS2	0.833	1.710			
PHS3	0.821	1.415			
Continuous use intention (CUI), Reflective			.703	0.860	0.756
CUI1	0.946	1.417			
CUI2	0.786	1.417			

[a]N/A: not applicable.

XSL•FO
RenderX

Table 4. Discriminant validity.[a]

Constructs	Continuous use intention	Information quality	Perceived health status	Perceived usefulness	Service quality	System quality	User satisfaction
Continuous use intention	*0.870*	—[b]	—	—	—	—	—
Information quality	0.489	—	—	—	—	—	—
Perceived health status	0.466	0.430	*0.827*	—	—	—	—
Perceived usefulness	0.578	0.618	0.458	*0.924*	—	—	—
Service quality	0.390	0.525	0.321	0.496	—	—	—
System quality	0.452	0.572	0.293	0.699	0.367	—	—
User satisfaction	0.574	0.514	0.513	0.714	0.426	0.478	*0.908*

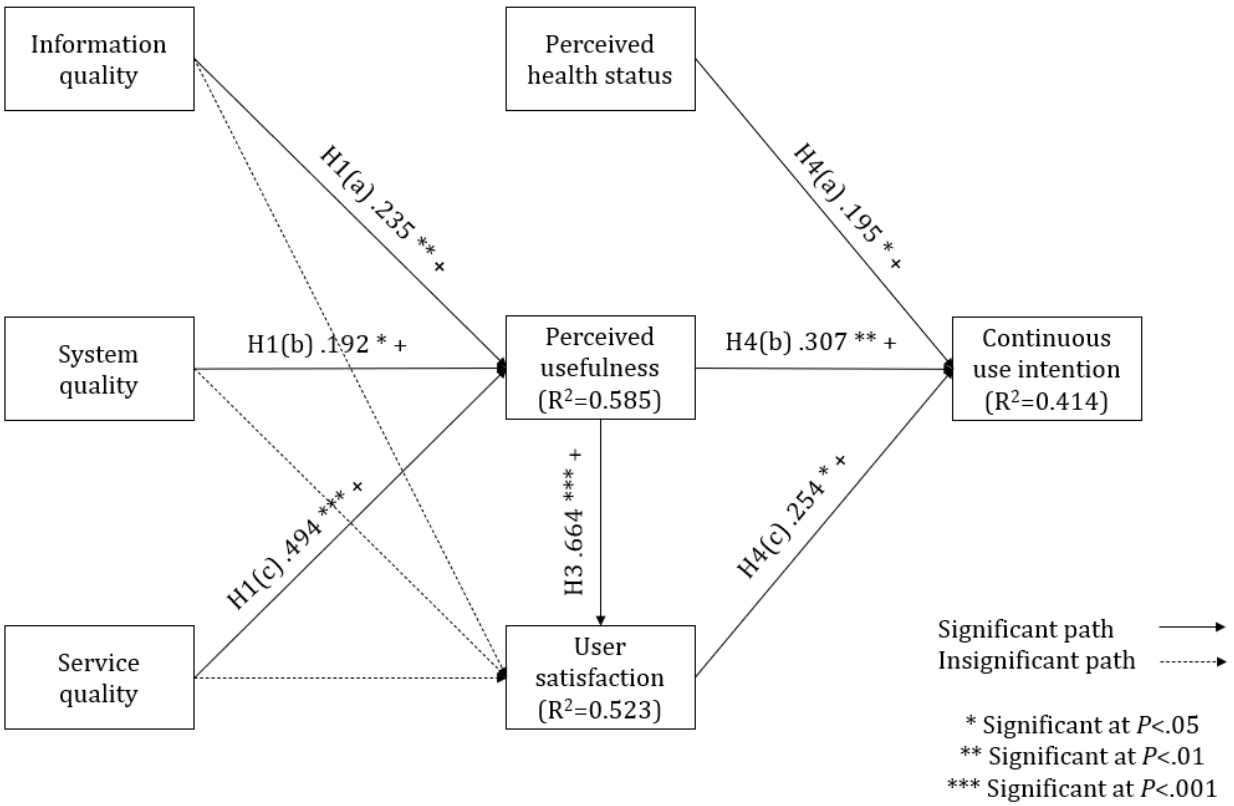
[a]The values presented in italics on the diagonal are the square roots of average variance extracted (the variance shared between the constructs and their measures). Off-diagonal values are the correlation coefficients for each construct in the relevant rows and columns.

[b]Not applicable.

Structural Model and Hypothesis Testing

For the path coefficient (β) and variance explained (R^2), our results showed that information quality, system quality, and service quality were all positively associated with perceived usefulness. Perceived usefulness and user satisfaction were positively associated with continuous use intention. Perceived usefulness was positively associated with user satisfaction. However, opposite to the original hypothesis, perceived health status was also positively associated with continuous use intention. Therefore, hypotheses 1(a), 1(b), 1(c), 3, 4(b), and 4(c) were confirmed but hypotheses 2(a), 2(b), 2(c), and 4(a) were not confirmed (Figure 2, Table 5, and Multimedia Appendix 2). The effects of service quality on perceived usefulness and the effects of perceived usefulness on user satisfaction were both large, while other paths showed small effects. The Q^2 values for all the 3 endogenous constructs were positive, that is, perceived usefulness (0.400), user satisfaction (0.337), and continuous use intention (0.144), which confirmed the model's predictive accuracy.

Figure 2. The validated theoretical model. "H" refers to the hypotheses in this study. +: positive association; –: negative association.



XSL•FO
RenderX

Table 5. Structural model validation and hypothesis testing results.

Path	Standard error	*t (df)*	*P* value	Effect size (f^2)	Hypothesis testing
H1(a): IQ→PU[a]	0.083	2.839 (9)	.005	Small (0.074)	Confirmed
H1(b): SysQ→PU[b]	0.084	2.293 (9)	.02	Small (0.064)	Confirmed
H1(c): SerQ→PU[c]	0.089	5.525 (9)	<.001	Large (0.391)	Confirmed
H2(a): IQ→US[d]	0.094	1.193 (9)	.23	None (0.014)	Not confirmed
H2(b): SysQ→US[e]	0.085	0.761 (9)	.45	None (0.006)	Not confirmed
H2(c): SerQ→US[f]	0.087	0.852 (9)	.39	None (0.005)	Not confirmed
H3: PU→US[g]	0.100	6.632 (9)	<.001	Large (0.384)	Confirmed
H4(a): PHS→CUI[h]	0.092	2.121 (9)	.03	Small (0.047)	Not confirmed
H4(b): PU→CUI[i]	0.107	2.879 (9)	.004	Small (0.077)	Confirmed
H4(c): US→CUI[j]	0.122	2.082 (9)	.04	Small (0.049)	Confirmed

[a]Hypothesis 1(a): Information quality is positively associated with patients' perceived usefulness of mobile health services.

[b]Hypothesis 1(b): System quality is positively associated with patients' perceived usefulness of mobile health services.

[c]Hypothesis 1(c): Service quality is positively associated with patients' perceived usefulness of mobile health services.

[d]Hypothesis 2(a): Information quality is positively associated with user satisfaction of mobile health services.

[e]Hypothesis 2(b): System quality is positively associated with user satisfaction of mobile health services.

[f]Hypothesis 2(c): Service quality is positively associated with user satisfaction of mobile health services.

[g]Hypothesis 3: Perceived usefulness of mobile health services is positively associated with user satisfaction with such services.

[h]Hypothesis 4(a): Perceived health status is negatively associated with continuous use intention of mobile health services.

[i]Hypothesis 4(b): Perceived usefulness is positively associated with continuous use intention of mobile health services.

[j]Hypothesis 4(c): User satisfaction is positively associated with continuous use intention of mobile health services.

Discussion

Principal Results

This empirical study proposes a theoretical model to predict and explain patients' intention to continue using mHealth services for self-management of chronic conditions. This model is derived from a hybrid model synthesized from the IS continuance model and IS success model [27,28]. The statistical assessment confirmed the reliability and validity of its measurement scale (Table 3 and Table 4). Six out of 10 original hypotheses about the relationships among 7 variables were confirmed (Table 5). Information quality, system quality, and service quality have a significant positive influence on perceived usefulness but not on user satisfaction. Perceived usefulness has a significant positive influence on user satisfaction. Perceived usefulness and user satisfaction have positive effects on participants' intention to continue using the mHealth services.

Contrary to the original hypothesis of a negative association, the patients' perceived health status is positively associated with their continuous use intention. This may be explained by the motivation effect of positive reward stemming from decreased blood pressure, which can directly motivate patients to form a virtuous cycle in self-management of chronic conditions, that is, the better the result in blood pressure control, the more actively a patient will use the app. Another possible reason is that frequent use of self-monitoring mechanisms, that is, entering and monitoring their own health data, has become a daily routine or habit. Thus, it is no longer influenced by intention, which only impacts voluntary use [60]. Lee et al [25] found that after introducing the self-monitoring function, the negative slope of the downward trend in mHealth service usage was alleviated in patients for self-management of their general health, which supported the long-term positive effects of self-monitoring for chronic conditions. However, according to the health belief model [61], which defines the key factors that influence health behaviors, one of the most important factors is the perceived severity, that is, belief of the consequence of the conditions. Therefore, once a patient's blood pressure is controlled, the person's perceived health status is improved and the perceived severity is relieved; therefore, the use intention will decrease. We consider that this paradox may be moderated by health literacy [15,62]. For patients with high health literacy, awareness of the negative consequences of not using the app coupled with the motivating effect brought by the perceived usefulness will motivate them to use the app continuously. Conversely, patients with low health literacy, owing to their lack of understanding of the negative effect of discontinued use, may not intend to use the app after a period or when their blood pressure is controlled. This is similar to the findings of Guo et al [23] that a patient's health consciousness has a significant positive impact on the relationship between social media influence and continuous use intention. Therefore, health literacy could be a moderator for the relationship between perceived health status and continued use intention, which can be further tested in a follow-up study.

XSL•FO
RenderX

The results also support the importance of providing feedback to patients about their health status that is reflected by the vital signs they uploaded in mHealth services [8]. Awareness of their own health status, particularly if the disease is deteriorating, will drive them to formulate an intention to continuously use the mHealth services. Conversely, when the health status is improving, patients might feel bored by repeatedly entering and uploading data day after day through the mHealth app, as reported by Biduski et al [63]. Without the feedback and awareness about their own health status, patients would assume they are in good health and gradually lose the intention to use the services.

Contributions of This Study

Theoretical Contribution

Our paper, for the first time, integrated 2 classic information system models for investigating the continuous use of mHealth services to support patient self-management of chronic conditions in the hypertension context [64]. The IS continuance model emphasizes the impact of expectation on perceived usefulness and user satisfaction based on the expectation-confirmation theory [27,29]. However, it does not clearly indicate which factors of the expectations are confirmed. The IS success model clarifies and supplements the 3 factors of the expectation: information quality, system quality, and service quality [28].

As hypothesized, our study confirms the cascading effects of 3 antecedent variables, that is, information quality, system quality, and service quality, on a patient's continuous use intention through the intermittent variables, that is, perceived usefulness and user satisfaction. Our findings are in line with Wu's findings that information quality and service quality have a significant positive impact on perceived usefulness [31]. In our study, service quality plays a more important role than information quality and system quality, which may indicate the importance of ongoing support from health care providers for patients to continuously use the mHealth app. A previous study in the same population showed that the relationship with the health care provider is one of the determinants for the patient's intention to use the same app [13]. Biduski et al [63] also found that the most satisfactory experience using mHealth services in self-management of chronic conditions mainly occurred in the first week and concentrated on the practicality of treatment monitoring. The service in Wu's study [31] was a web-based health community. Patients usually communicate with their doctors and seek relevant information through a web-based community. Our service provides patients with a disease-focused intervention specifically for the management of hypertension. In addition to obtaining information, the patients also require ongoing support provided by a health manager, which is an important component of service. This is also reflected in the positive evaluation of patients in questions 6 and 7, that is, "The health manager can resolve problems which I encountered when using the Blood Pressure Assistant" (4.3/5) and "The health manager gives me individual guidance in the follow-up phone calls" (4.1/5). The services provided by the health manager include guidance and assurance on self-management when abnormal blood pressure levels are detected. The health manager would call the patients or their family to discuss and modify the management plan. These interactions give patients the comfort that they receive full attention and that they receive high-quality service from their health care providers. These interactions enhance their rapport with health care providers and trust and recognition of service quality so that patients can perceive the usefulness of mHealth and generate the intention to make continuous use of the app. Our finding confirms the vital role that health care providers play in ensuring that patients use the mHealth service continuously.

The finding that the system quality has less influence on perceived usefulness in comparison with information and service quality may be explained by the complexity and persistent effort required for the management of chronic conditions [65,66]. Unlike using other systems, patients are more concerned about whether the service and information provided by the app can help them control the disease. This may also explain why user satisfaction is not directly affected by information quality, service quality, and system quality but indirectly through perceived usefulness.

Practical Contribution

Although the rapid development of mHealth services and the initial acceptance by patients have brought new opportunities for managing chronic conditions, owing to the lasting nature of these conditions and the long-term requirements for behavior modification, only patients who continue to use the mHealth service can benefit from such services. Our research results show that 3 independent factors, namely, information quality, system quality, and service quality determine patients' intention to continue to use mHealth services through the mediation of perceived usefulness and user satisfaction. In the context of this study, service quality (the assistance and feedback of health care providers), information quality (the provision of reliable and relevant information), and system quality (an easy-to-learn and easy-to-operate system) are essential for patients' continuous use of an mHealth service. Therefore, to successfully introduce mHealth innovation into self-management of chronic conditions, it is necessary to focus on improving the responsibility and ability of health care providers and to continue to provide and update the educational information for patients to improve their awareness of the disease threat and effective self-management methods. These strategies will further enable patients to understand the usefulness of mHealth services and manage their chronic conditions by using the services continuously.

Limitations and Future Work

The first limitation of this study is the limited geographical and social coverage of the study population, which is limited to an underdeveloped area in China. The sample size is relatively small, with only 129 patients, despite reaching the threshold for theoretical sampling. The mHealth app is purposely built to test the feasibility of hypertension control by using an app; thus, its usability may not be representative of all mHealth apps for any type of chronic conditions. Thus, those implementing the findings should be cautious about generalizing these findings. Second, the latent variables were selected based on the group's previous research experience and literature review; other factors may also need to be captured in the model that we tested, for

XSL•FO
RenderX

example, health literacy. Third, as a trade-off for avoiding survey fatigue and improving response rates [28], many constructs were measured by 2 items instead of 3 items suggested by Hair et al [39], which were suboptimal from the perspective of measurement theory. Since the questionnaire was answered voluntarily by the users of the mHealth service, the results cannot avoid the positive bias in sampling, that is, nonusers would not answer the questionnaires. In addition, no control or moderating variables, that is, patient characteristics, were included in the structural model for the sake of meeting the threshold number of measures in SEM. Future research can investigate the impact of the control or moderating variables, that is, perceived risk and patient-doctor relationship [13,15] on perceived usefulness, user satisfaction, or continuous use intention. Future research needs to also be conducted as an empirical field study on the long-term effects of mHealth services in natural settings to derive generalizable insight for improving the practice of implementing mHealth services.

Conclusions

This study developed a model and questionnaire survey instrument to measure the success of patients' continuous use of mHealth services for self-management of chronic conditions. This study shows that patients' intention to continue to use an mHealth service for self-management of chronic conditions is influenced by information quality, system quality, and service quality, through the mediation of perceived usefulness and user satisfaction. The patients' perceived health status also has a significant positive influence on their continuous use intention. The validated model and measurement scale are useful for the routine evaluation of patients' continuous use of mHealth services, which is also important for evaluating the operational effect of the mHealth program. The research model and the questionnaire survey instrument developed can be used for routine identification of the areas of mHealth management provided to the patients that support their use of the mHealth services that need improvement, that is, information quality, mobile app usability, and technical and health care provider services. These findings also enrich the body of knowledge of continuous use of mHealth for self-management of chronic conditions. Future research can apply the model and questionnaire survey instrument to other types of mHealth services.

Acknowledgments

The authors would like to thank all the participants for their responses to the questionnaire. Juhong Chen and Yinping Li are acknowledged for their assistance in content validation of the questionnaire survey instrument. Zheyu Wang, Hui Lin, and Kaili Dou are also acknowledged and appreciated for their assistance in data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Cross-loading of the latent variables.
[PDF File (Adobe PDF File), 132 KB - jmir_v23i7e26670_app1.pdf]

Multimedia Appendix 2
The direct, indirect, and total effects of antecedent and dependent variables on the other dependent variables.
[PDF File (Adobe PDF File), 103 KB - jmir_v23i7e26670_app2.pdf]

References

1. Debon R, Coleone JD, Bellei EA, De Marchi ACB. Mobile health applications for chronic diseases: A systematic review of features for lifestyle improvement. Diabetes Metab Syndr 2019;13(4):2507-2512. [doi: 10.1016/j.dsx.2019.07.016] [Medline: 31405669]
2. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]
3. Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
4. Song T, Yu P. mHealth. In: Gu D, Dupre ME, editors. Encyclopedia of Gerontology and Population Aging. Cham: Springer International Publishing; 2020:1-7.
5. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
6. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045 [FREE Full text] [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]

XSL•FO
RenderX

7. Alsharo M, Alnsour Y, Alabdallah M. How habit affects continuous use: evidence from Jordan's national health information system. Inform Health Soc Care 2020 Jan;45(1):43-56. [doi: 10.1080/17538157.2018.1540423] [Medline: 30457025]
8. Song T, Liu F, Deng N, Qian S, Cui T, Guan Y, et al. A Comprehensive 6A Framework for Improving Patient Self-Management of Hypertension Using mHealth Services: Qualitative Thematic Analysis. J Med Internet Res 2021 Jun 21;23(6):e25522 [FREE Full text] [doi: 10.2196/25522] [Medline: 34152272]
9. Perez S. Nearly 1 in 4 people abandon mobile apps after only one use. TechCrunch. URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/#:~:text=Apple's%20iTunes%20App%20Store%20is,devices%20is%20still%20quite%20small [accessed 2020-12-06]
10. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
11. Savoli A, Barki H, Pare G. Examining How Chronically Ill Patients' Reactions to and Effective Use of Information Technology Can Influence How Well They Self-Manage Their Illness. MISQ 2020 Jan 01;44(1):351-389. [doi: 10.25300/misq/2020/15103]
12. Burton-Jones A, Volkoff O. How Can We Develop Contextualized Theories of Effective Use? A Demonstration in the Context of Community-Care Electronic Health Records. Information Systems Research 2017 Sep;28(3):468-489. [doi: 10.1287/isre.2017.0702]
13. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' Acceptance of Smartphone Health Technology for Chronic Disease Management: A Theoretical Model and Empirical Test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
14. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
15. Zhu Z, Liu Y, Che X, Chen X. Moderating factors influencing adoption of a mobile chronic disease management system in China. Inform Health Soc Care 2018 Jan;43(1):22-41. [doi: 10.1080/17538157.2016.1255631] [Medline: 28068149]
16. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
17. Zhou L, Bao J, Parmanto B. Systematic Review Protocol to Assess the Effectiveness of Usability Questionnaires in mHealth App Studies. JMIR Res Protoc 2017 Aug 01;6(8):e151 [FREE Full text] [doi: 10.2196/resprot.7826] [Medline: 28765101]
18. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth App for Self-Management of Chronic Lower Back Pain (Limbr): Pilot Study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]
19. Torbjørnsen A, Småstuen MC, Jenum AK, Årsand E, Ribu L. Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]
20. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]
21. Loh KP, Ramsdale E, Culakova E, Mendler JH, Liesveld JL, O'Dwyer KM, et al. Novel mHealth App to Deliver Geriatric Assessment-Driven Interventions for Older Adults With Cancer: Pilot Feasibility and Usability Study. JMIR Cancer 2018 Oct 29;4(2):e10296 [FREE Full text] [doi: 10.2196/10296] [Medline: 30373733]
22. Thirumalai M, Rimmer JH, Johnson G, Wilroy J, Young H, Mehta T, et al. TEAMS (Tele-Exercise and Multiple Sclerosis), a Tailored Telerehabilitation mHealth App: Participant-Centered Development and Usability Study. JMIR Mhealth Uhealth 2018 May 24;6(5):e10181 [FREE Full text] [doi: 10.2196/10181] [Medline: 29798832]
23. Guo X, Chen S, Zhang X, Ju X, Wang X. Exploring Patients' Intentions for Continuous Usage of mHealth Services: Elaboration-Likelihood Perspective Study. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17258 [FREE Full text] [doi: 10.2196/17258] [Medline: 32250277]
24. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
25. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
26. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
27. Bhattacherjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
28. Delone WH, McLean R. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. Journal of Management Information Systems 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
29. Oliver RL. A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research 1980 Nov;17(4):460. [doi: 10.2307/3150499]

XSL•FO
RenderX

30. Yu P, Qian S. Developing a theoretical model and questionnaire survey instrument to measure the success of electronic health records in residential aged care. PLoS One 2018;13(1):e0190749 [FREE Full text] [doi: 10.1371/journal.pone.0190749] [Medline: 29315323]
31. Wu B. Patient Continued Use of Online Health Care Communities: Web Mining of Patient-Doctor Communication. J Med Internet Res 2018 Apr 16;20(4):e126 [FREE Full text] [doi: 10.2196/jmir.9127] [Medline: 29661747]
32. Zheng Y, Zhao K, Stylianou A. The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation. Decision Support Systems 2013 Dec;56:513-524. [doi: 10.1016/j.dss.2012.11.008]
33. Alsabawy A, Cater-Steel A, Soar J. Determinants of perceived usefulness of e-learning systems. Computers in Human Behavior 2016 Nov;64(1):843-858. [doi: 10.1016/j.chb.2016.07.065]
34. Watts S, Shankaranarayanan G, Even A. Data quality assessment in context: A cognitive perspective. Decision Support Systems 2009 Dec;48(1):202-211. [doi: 10.1016/j.dss.2009.07.012]
35. Gelderman M. The relation between user satisfaction, usage of information systems and performance. Information & Management 1998 Aug;34(1):11-18. [doi: 10.1016/s0378-7206(98)00044-5]
36. Kim Y, Lee HS. Quality, perceived usefulness, user satisfaction, and intention to use: An empirical study of ubiquitous personal robot service. Asian Social Science 2014 May 30;10(11):1-16. [doi: 10.5539/ass.v10n11p1]
37. Kim GM, Hong MS, Noh W. Factors affecting the health-related quality of life in community-dwelling elderly people. Public Health Nurs 2018 Nov;35(6):482-489. [doi: 10.1111/phn.12530] [Medline: 29947059]
38. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
39. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. EBR 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
40. Parry E, Farndale E, Brewster C, Morley MJ. Balancing Rigour and Relevance: The Case for Methodological Pragmatism in Conducting Large‐Scale, Multi‐country and Comparative Management Studies. Brit J Manage 2020 Apr 27;32(2):273-282. [doi: 10.1111/1467-8551.12405]
41. Bhattacherjee A, Perols J, Sanford C. Information Technology Continuance: A Theoretic Extension and Empirical Test. Journal of Computer Information Systems 2015 Dec 11;49(1):17-26. [doi: 10.1080/08874417.2008.11645302]
42. Bhattacherjee, Premkumar. Understanding Changes in Belief and Attitude toward Information Technology Usage: A Theoretical Model and Longitudinal Test. MIS Quarterly 2004;28(2):229. [doi: 10.2307/25148634]
43. Karakas SA. Investıgatıng alexıthymıa and socıal support in patıents wıth hypertensıon and asthma. J Psy Nurs 2016 Dec;7(2):68-74. [doi: 10.5505/phd.2016.40085]
44. Kuo SZ, Haftek M, Lai JC. Factors Associated with Medication Non-adherence in Patients with End-Stage Liver Disease. Dig Dis Sci 2017 Feb;62(2):543-549 [FREE Full text] [doi: 10.1007/s10620-016-4391-z] [Medline: 27933471]
45. Hair JJ, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Newbury Park, California: SAGE Publications Inc; Apr 05, 2016:1-384.
46. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
47. Wondola D, Aulele S, Lembang F. Partial least square (PLS) method of addressing multicollinearity problems in multiple linear regressions (case studies: cost of electricity bills and factors affecting it). 2020 Feb 27 Presented at: The 5th International Conference on Basic Sciences; 5-6 September 2019; Kota Ambon, Maluku. [doi: 10.1088/1742-6596/1463/1/012006]
48. Reinartz W, Haenlein M, Henseler J. An empirical comparison of the efficacy of covariance-based and variance-based SEM. International Journal of Research in Marketing 2009 Dec;26(4):332-344. [doi: 10.1016/j.ijresmar.2009.08.001]
49. Urbach N, Ahlemann F. Structural equation modeling in information systems research using partial least squares. JITTA. 2010. URL: https://aisel.aisnet.org/jitta/vol11/iss2/2/ [accessed 2021-06-28]
50. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
51. Götz O, Liehr-Gobbers K, Krafft M. Evaluation of structural equation models using the partial least squares (PLS) approach. In: Handbook of Partial Least Squares. Berlin, Heidelberg: Springer; 2010:691-711.
52. Shmueli, Koppius. Predictive Analytics in Information Systems Research. MIS Quarterly 2011;35(3):553. [doi: 10.2307/23042796]
53. Rigdon EE. Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. Long Range Planning 2012 Oct;45(5-6):341-358. [doi: 10.1016/j.lrp.2012.09.010]
54. Nitzl C. The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development. Journal of Accounting Literature 2016 Dec;37:19-35. [doi: 10.1016/j.acclit.2016.09.003]
55. Geisser S. A predictive approach to the random effect model. Biometrika 1974;61(1):101-107. [doi: 10.1093/biomet/61.1.101]
56. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;36(2):111-133. [doi: 10.1111/j.2517-6161.1974.tb00994.x]

XSL•FO
RenderX

57. Bentler PM, Chou C. Practical Issues in Structural Modeling. Sociological Methods & Research 2016 Jun 30;16(1):78-117. [doi: 10.1177/0049124187016001004]
58. Becker J, Ringle CM, Sarstedt M, Völckner F. How collinearity affects mixture regression results. Mark Lett 2014 May 7;26(4):643-659. [doi: 10.1007/s11002-014-9299-9]
59. Franke G, Sarstedt M. Heuristics versus statistics in discriminant validity testing: a comparison of four procedures. INTR 2019 Jun 03;29(3):430-447. [doi: 10.1108/intr-12-2017-0515]
60. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
61. Jones CJ, Smith H, Llewellyn C. Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review. Health Psychol Rev 2014;8(3):253-269. [doi: 10.1080/17437199.2013.802623] [Medline: 25053213]
62. Wehmeyer MMH, Corwin CL, Guthmiller JM, Lee JY. The impact of oral health literacy on periodontal health status. J Public Health Dent 2014;74(1):80-87 [FREE Full text] [doi: 10.1111/j.1752-7325.2012.00375.x] [Medline: 23121152]
63. Biduski D, Bellei EA, Rodriguez JPM, Zaina LAM, De Marchi ACB. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. Computers in Human Behavior 2020 Mar;104:106169. [doi: 10.1016/j.chb.2019.106169]
64. Song T, Qian S, Cui T, Yu P. The Use of Theory in Mobile Health Interventions for Patient Self-Management of Chronic Diseases. Stud Health Technol Inform 2019 Aug 21;264:1982-1983. [doi: 10.3233/SHTI190745] [Medline: 31438439]
65. Bossen C, Jensen LG, Udsen FW. Evaluation of a comprehensive EHR based on the DeLone and McLean model for IS success: approach, results, and success factors. Int J Med Inform 2013 Oct;82(10):940-953. [doi: 10.1016/j.ijmedinf.2013.05.010] [Medline: 23827768]
66. Guo J. Smartphone-Powered Electrochemical Biosensing Dongle for Emerging Medical IoTs Application. IEEE Trans. Ind. Inf 2018 Jun;14(6):2592-2597. [doi: 10.1109/tii.2017.2777145]

Abbreviations

IS: information systems
mHealth: mobile health
PLS-SEM: partial least squares structural equation modelling
VIF: variance inflation factor

Edited by R Kukafka; submitted 21.12.20; peer-reviewed by E Bellei, A Wahbeh, B Nanda, P Tripathi; comments to author 01.02.21; revised version received 18.02.21; accepted 06.05.21; published 13.07.21.

Please cite as:
Song T, Deng N, Cui T, Qian S, Liu F, Guan Y, Yu P
Measuring Success of Patients' Continuous Use of Mobile Health Services for Self-management of Chronic Conditions: Model Development and Validation
J Med Internet Res 2021;23(7):e26670
URL: https://www.jmir.org/2021/7/e26670
doi:10.2196/26670
PMID:34255685

©Ting Song, Ning Deng, Tingru Cui, Siyu Qian, Fang Liu, Yingping Guan, Ping Yu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effects of Continuous Usage of a Diabetes Management App on Glycemic Control in Real-world Clinical Practice: Retrospective Analysis

Yu-Zhen Tu[1], MSc; Ya-Ting Chang[1], BSc; Hung-Yi Chiou[2], PhD; Ken Lai[1], MSc

[1]H2 Inc, Taipei, Taiwan

[2]School of Public Health, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ken Lai, MSc
H2 Inc
4F, No 32, Aly 18, Ln 478
Ruiguang Rd, Neihu Dist
Taipei
Taiwan
Phone: 886 287976661 ext 105
Email: klai@health2sync.com

Abstract

Background: The efficacy of digital technology in improving diabetes management has typically been demonstrated through studies such as randomized controlled trials, which have reported a steeper reduction in hemoglobin A_{1c} (HbA_{1c}) values for patients who adopted a digital solution. However, evidence from real-world clinical practice is still limited.

Objective: This study aimed to evaluate the effectiveness of digital interventions by tracking HbA_{1c} improvements over 1 year in real-world clinical settings.

Methods: Patients used the Health2Sync mobile app to track self-measured outcomes and communicate with health care professionals (HCPs). HCPs used the web-based Patient Management Platform to monitor patient data, view test results from clinical laboratories, and communicate with patients. Patients who have been onboarded for at least 13 months and have consecutive HbA_{1c} findings for 5 quarters were included in the analysis. They were then stratified into 3 groups (high, mid, and low retention) based on their level of use of Health2Sync in the first 6 months of onboarding. A mixed model was built to compare the slopes of the rate of reduction in HbA_{1c} among the groups. In addition, these patients' retention on the app from the seventh to the 12th month was verified through multiple comparisons.

Results: A sample of 2036 users was included in the analysis. With the mixed model coefficient estimates, we found that app users had significant HbA_{1c} percentage reductions as the passed quarter count increased (t=–9.869; P<.001), and that effectiveness increased in the high (t=–5.173) and mid retention (t=–6.620) groups as the interaction effects were significantly negative compared to that in the low retention group (P<.001) in the passed quarter count. The low retention group also had the highest average HbA_{1c} value at the end of 13 months (high: 7.01%, SD 1.02%; mid: 6.99%, SD 1.00%; low: 7.17%, SD 1.14%) (Bonferroni correction: high vs low, P=.07; mid vs low, P=.02; high vs mid, P>.99). The level of use of the app remained consistent in the seventh to the 12th month after onboarding (high: 5.23 [SD 1.37] months, mid: 2.43 [SD 1.68] months, low: 0.41 [SD 0.97] months) (P<.001).

Conclusions: Our analysis shows that continuous usage of the diabetes management app is associated with better glycemic control in real-world clinical practice. Further studies are required to reveal the efficacy for specific diabetes types and to observe effects beyond 1 year.

(J Med Internet Res 2021;23(7):e23227) doi:10.2196/23227

KEYWORDS

app; diabetes care; diabetes; digital intervention; digital therapeutics; glycemic control; mobile app; mHealth; real-world data; therapy

XSL•FO
RenderX

Introduction

Global diabetes prevalence in the adult population is estimated to have grown from 8.5% in 2014 to 9.3% in 2019 and is projected to be over 10% by 2030 [1,2]. In Taiwan, the prevalence of diabetes approached 9.8% in 2016 [3]. Caring for patients with diabetes and a couple of its associated complications—chronic kidney disease and acute kidney failure—costed more than US $2.5 billion in Taiwan in 2019, which accounted for 11.4% of the total health care expenditure in that year [4]. As the cost of diabetes care continues to increase in the foreseeable future, health care providers will need to consider adopting digital solutions to manage patients with diabetes in an effective and scalable manner.

Increasing evidence suggests that interventions with digital technology enhance the effect of conventional care practices on patients with diabetes. For example, patients who received coaching and decision support from the WellDoc app presented greater reductions in hemoglobin A_{1c} (HbA_{1c}) levels compared to those in the control group [5,6]. A cohort study using the FareWell app suggested the same effect on adults with type 2 diabetes, and the users reported greater confidence in managing the disorder [7]. Furthermore, multiple meta-analyses on randomized controlled trials (RCTs) have shown that interventions including mobile apps help patients, especially those with type 2 diabetes, to lower their HbA_{1c} levels significantly without notable adverse effects [8-14]. These apps usually allow patients with diabetes to log and visualize their self-monitored data, provide education and feedback, or provide communication channels between patients and their caregivers or peers [12,13,15]. From the patient's perspective, these features were also deemed as characteristics of a great app [16].

Although the aforementioned mobile apps are intended to be used by patients in their daily lives, few studies have verified the effects of mobile apps with real-world data. A previous study [17] using the mobile version of the One Drop diabetes management platform reported a substantial reduction in HbA_{1c} values of users for the first 2 entries in the app regardless of diabetes type, but the study did not compare effects between app users and nonusers and did not examine the long-term effects of app engagement [17].

This study aims to address the aforementioned issues through a retrospective analysis. We evaluated the difference in HbA_{1c} improvement between active app users and those who drop out from the app in a real-world context, and monitored their glycemic control status for 1 year.

Methods

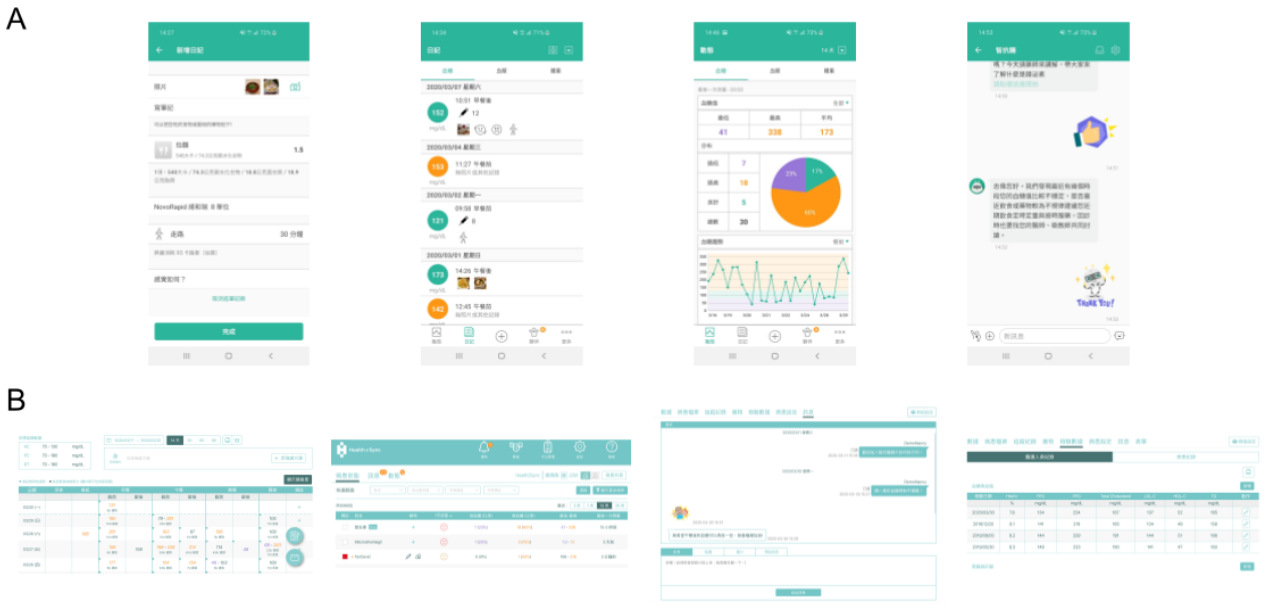
Health2Sync App and Patient Management Platform

The Health2Sync mobile app and web-based Patient Management Platform were launched in 2014 to support patients with chronic diseases such as diabetes. The Health2Sync solution was developed to support patients to make behavioral changes through a do-track-learn cycle and enable those around the patient to care for the patient. The mobile app, intended for patients, is free to download for iOS and Android users worldwide. The Patient Management Platform is designed for health care professionals (HCPs) to care for patients in clinics or hospitals remotely.

The Health2Sync app has numerous features that support diabetes management (Figure 1A). To start with, users can record their self-measured outcomes manually or by synchronizing with glucose meters, sphygmomanometers, and weight scales. They can also log their daily behaviors such as diet, medication, and exercise. Users can review their past records to remind themselves of how their behaviors are linked to the outcomes. On the dashboard, users can visualize their recent progress and trends by referring to simple charts and statistics. Health2Sync's most differentiating feature is its "Partners" function, which enables users to connect with HCPs, family members, or peers so that they can collaborate with the patient to manage his/her condition. The "Partners" feature allows hospitals or clinics that use the Patient Management Platform to view the patient's data and communicate with the patient through a messaging feature on the platform. Every user of the app also has the Health2Sync bot as a default partner that provides automated analyses, alerts, encouragements, and educational content. In addition, the app has a peer functionality that lets users learn and interact with other patients with diabetes who use the app.

XSL•FO
RenderX

Figure 1. Screenshots of the Health2Sync app and Patient Management Platform.



The recorded outcomes and behavioral data from patients will be sent in real time to the Patient Management Platform to be reviewed by HCPs (Figure 1B). On the platform, HCPs can also send messages to patients to provide care and answer questions. From the Patient Management Platform's dashboard, HCPs can quickly scan their pool of patients to check their latest status and identify any patients requiring immediate care. The alert functionality will notify HCPs of critical patient trends or events. The Patient Management Platform can also be integrated with clinical laboratories so that HCPs can view the results of laboratory tests directly on the platform. Test results including those for HbA_{1c}, blood glucose, lipid profile, and renal and liver function are supported.

Data Collection

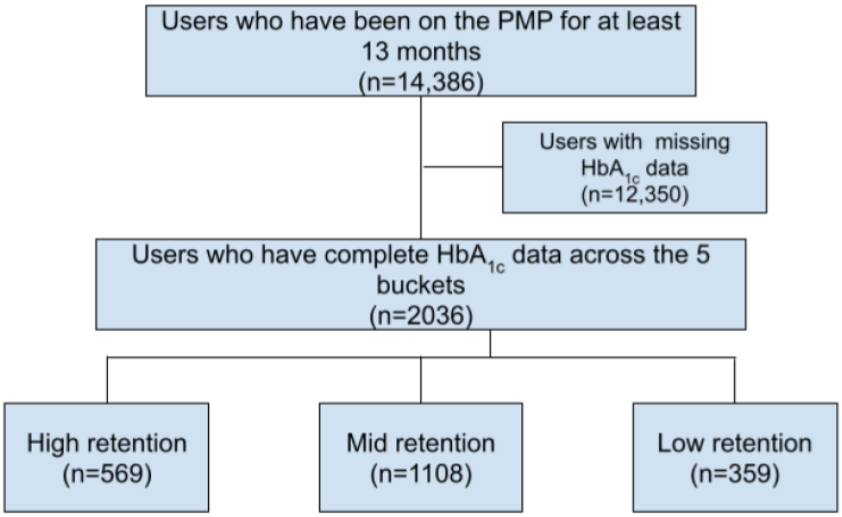
Users' HbA_{1c} records on the Patient Management Platform were included. All the HbA_{1c} values were either inputted by HCPs or synchronized from clinical laboratories. To assess the effects of the intervention at different time points, we looked at when each patient added a clinic or hospital as a partner in Health2Sync and determined how many months after adding the partner the HbA_{1c} values were recorded. The calculation method considered the difference in the number of days between the date when users added the partner and that of each HbA_{1c} recording, divided by 30. According to the calculated numbers, these records were then categorized into five time buckets from B_0 to B_4: [–2, 1), [1, 4), [4, 7), [7, 10), and [10, 13). The buckets were designed to reflect the common practice of determining HbA_{1c} levels once every 3 months in Taiwan, and that the first HbA_{1c} record can be in –2~1 month from the time when the partnership was created. We believe that as glycated hemoglobin levels typically reflect blood glucose values in the past 8-12 weeks [18], HbA_{1c} levels in the first month of the partnership would only be minimally affected by the intervention.

Subjects

Health2Sync users who were connected to clinics or hospitals via the Health2Sync platform for at least 13 months were included, thus yielding an initial pool of 14,386 users. Those with missing data in any of the 5-bucket HbA_{1c} observation periods were excluded owing to the requirement of a complete data set for subsequent modeling. At this stage, we excluded 12,350 users and retained 2036 users. These users were then separated into 3 groups in accordance with their app retention. In the first 6 months after adding the partner, users who opened the Health2Sync app every month were categorized as the "high retention" group, those who opened the app in only the first month were categorized as the "low retention" group, and the remaining patients were categorized as the "mid retention" group. Eventually, there were 569, 1108, and 359 users in the high, mid, and low retention groups, respectively. Demographic data of the included 2036 users and their distribution across the 3 retention groups are depicted in Table 1, although the majority of these included users preferred not to disclose their demographic status. Throughout this study, all users continued to receive diabetes care from the same clinic or hospital, otherwise there would be no HbA_{1c} records on the Patient Management Platform. Figure 2 shows the inclusion flow chart described above.

XSL•FO
RenderX

Figure 2. Inclusion flow chart of this study. The final 3 groups were separated by their retention in the first 6 months. Sample sizes of each stage are noted. PMP: Patient Management Platform.



Analyses

All the analyses in this study were performed with R (version 3.6.1, The R Foundation) [19].

User Characteristics

All users filled an onboarding form in the Health2Sync app where they self-reported their age, gender, and diabetes type. One-way analysis of variance (ANOVA) and the Pearson chi-square test were used for continuous and categorical variables, respectively, to assess the homogeneity of demographics across the 3 retention groups.

HbA_{1c} Changes

Mean (SD) HbA_{1c} levels of each group in B_0 and B_4 were calculated to compare the glycemic control status by app retention levels. One-way ANOVA was used to examine the heterogeneity across groups, and the pairwise *t* test with Bonferroni correction was applied for post hoc analysis.

We used the rate of change in HbA_{1c} levels, generated using the formula (HbA_{1c} value - mean HbA_{1c} in B_0) / mean HbA_{1c} in B_0, to assess the improvement in glycemic status across the groups. The rate of change was used instead of the absolute magnitude of change because users who had higher HbA_{1c} levels in B_0 also had more room for a reduction in its value; hence, the comparison of changes in absolute HbA_{1c} values across groups would be biased. In addition, considering that the reduction in HbA_{1c} levels was nonlinear with time, all the changes in rates were added by a constant so that the minimum value is 1, and all these values were log-transformed. These log-transformed values were incorporated in the mixed model as the dependent variable, where fixed effects included the time bucket numbers (ie, quarters passed, from 0 to 4), app retention groups, and their interaction. Each user's intercept was controlled in the random effects model [20,21].

Monitoring of App Retention

To follow the users' app retention in the subsequent 6 months (ie, months 7-12 after partnership creation), we enumerated the months after which they had opened the app. The purpose of this analysis is to verify whether these 3 groups of users maintained their usage behavior in the following months of the retrospective observation period. The differences across these groups were assessed using 1-way ANOVA, and a pairwise *t* test with Bonferroni correction was used for post hoc analysis.

Results

Table 1 presents user characteristics stratified by the 3 user groups based on different levels of app retention. Users were not required to provide their demographic profiles; hence, sample sizes may differ across demographic characteristics. We found no significant difference in age ($F_{2,806}$=1.441; *P*=.24) and gender (χ^2_{22}=0.3637; *P*=.83) distributions across the 3 groups; however, the diabetes type distribution was not even in these groups (χ^2_{24}=27.489; *P*<.001).

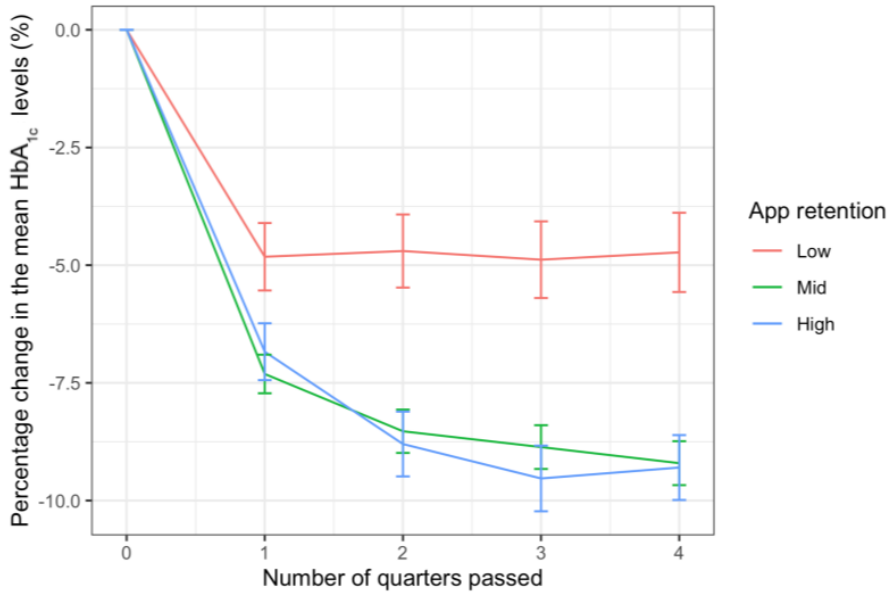
XSL•FO
RenderX

Table 1. User characteristics stratified by app retention (N=2036).

Characteristics	Total	High retention (n=569)	Mid retention (n=1108)	Low retention (n=359)	*P* value
Age (years), mean (SD); n	57.7 (13.4); 809	57.1 (12.3); 297	58.7 (13.7); 327	57.1 (14.4); 185	.24
Gender, n					.83
Male	506	226	175	105	
Female	461	214	157	90	
Diabetes type, n					<.001
Type 1	53	21	10	22	
Type 2	859	384	311	164	
Others	33	24	6	3	
HbA_{1c} in time bucket B_0 (%), mean (SD)	7.90 (1.74)	7.99 (1.86)	7.92 (1.72)	7.70 (1.60)	.04
HbA_{1c} in time bucket B_4 (%), mean (SD)	7.03 (1.03)	7.01 (1.02)	6.99 (1.00)	7.17 (1.14)	.02

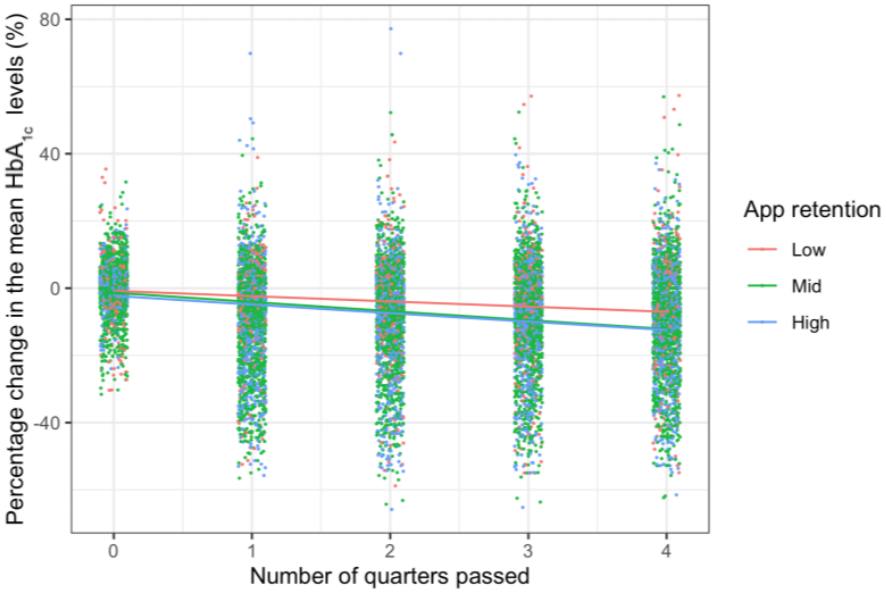
Users with higher app retention presented a greater reduction in the rate of change in HbA_{1c} levels (Figure 3). The group with the lowest retention presented the smallest reduction in HbA_{1c} levels in time bucket B_1 (–4.8%, SD 13.6%) and remained constant afterward up to time bucket B_4 (–4.7%, SD 15.9%), while the other groups presented a steeper reduction in time bucket B_1 (mid: –7.3%, SD 13.6%; high: –6.8%, SD 14.4%), and the decreasing trends continued up to time bucket B_4 (mid: –9.2%, 15.5%; high: –9.3%, SD 16.4%). The mixed model further verified the significance of this effect (Figure 4). Users presented significant HbA_{1c} percentage reductions as the passed quarter count increased (β estimate=$–9.626\times10^{-3}$; t=–9.869; P<.001). Furthermore, being in mid and high retention groups further augmented this reduction with time as the interactions were significantly negative when the low retention group was set as the baseline for the past quarter (mid: β estimate=$–7.360\times10^{-3}$; t=–6.620; P<.001; high: β estimate=$–6.419\times10^{-3}$; t=–5.173; P<.001). When the high retention group was set as the baseline for comparison, no significant difference was found between the augmentation effects between the mid and high retention groups for the past quarter (β estimate=$–9.410\times10^{-4}$; t=–1.007; P=.31). No main effect was observed for the retention groups when the low retention group was set as the baseline for comparison (mid: β estimate=$–3.294\times10^{-3}$; t=–0.691; P=.49; high: β estimate=$–8.951\times10^{-3}$; t=–1.687; P=.09).

Figure 3. Users' averaged percentage change in HbA_{1c} levels in each time bucket. The error bars represent the SE among users in the group.



XSL•FO
RenderX

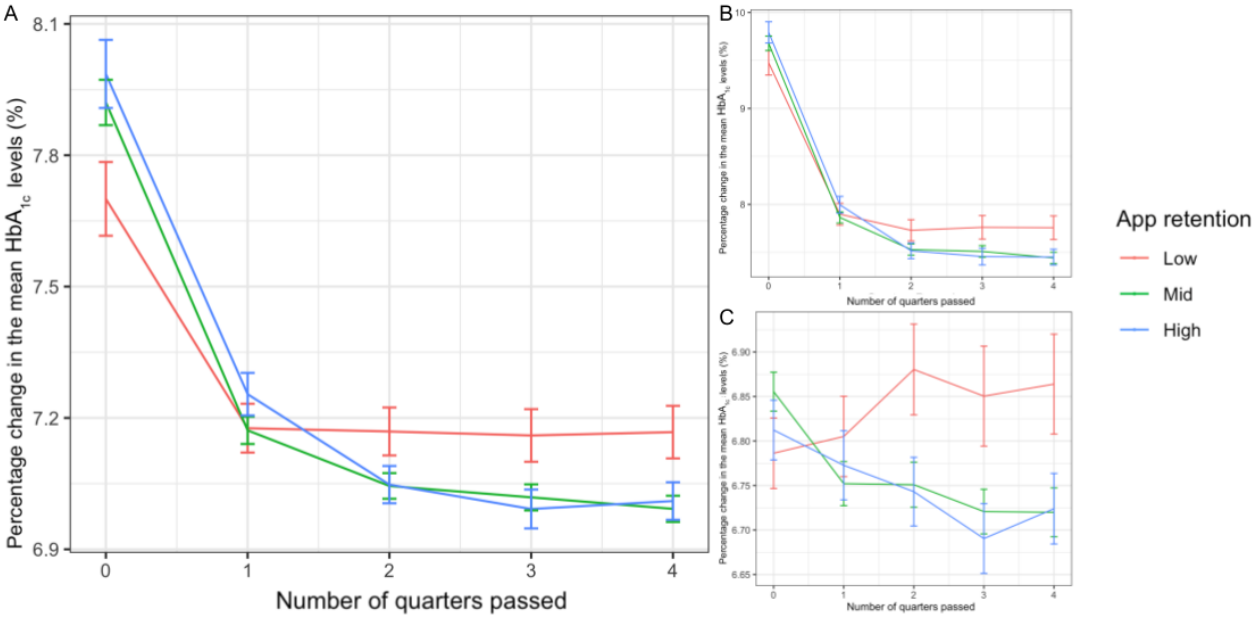
Figure 4. Jittered scatter plot depicting the relationship between HbA_{1c} percentage changes and time periods after joining the Patient Management Platform. Each dot represents 1 user's data at the time. The overlaid regression lines are based on the estimated coefficients from the mixed model.



The high and mid retention groups also presented a greater reduction in raw HbA_{1c} values and lower HbA_{1c} levels at the end of the observation period, although their raw HbA_{1c} values were higher in time bucket B_0 (Table 1 and Figure 5A). We observed significant differences in HbA_{1c} levels between the high and low retention groups in time bucket B_0 (high: 7.99%, SD 1.86%; 95% CI 7.83%-8.14%; mid: 7.92%, SD 1.72%; 95% CI 7.82%-8.02%; low: 7.70%, SD 1.60%; 95% CI 7.54%-7.87%) (Bonferroni correction: high vs low, *P*=.045; mid vs low, *P*=.11; high vs mid, *P*>.99) and between mid and low retention groups in time bucket B_4 (high: 7.01%, SD 1.02%; 95% CI 6.93%-7.09%; mid: 6.99%, SD 1.00%; 95% CI 6.93%-7.05%; low: 7.17%, SD 1.14%; 95% CI 7.05%-7.29%) (Bonferroni correction: high vs low, *P*=.07; mid vs low, *P*=.02; high vs mid, *P*>.99). In a stratified analysis, users with HbA_{1c} values of ≥8% in time bucket B_0 presented no significant difference in glycemic status at the beginning (high: 9.79%, SD 1.66%; 95% CI 9.58%-10.01%; mid: 9.68%, SD 1.54%; 95% CI 9.53%-9.83%; low: 9.48%, SD 1.41%; 95% CI 9.23%-9.73%) (F_2=1.632; *P*=.20), but the difference in HbA_{1c} levels between the mid and low retention groups in time bucket B_4 was still significant (high: 7.45%, SD 1.23%; 95% CI 7.29%-7.61%; mid: 7.44%, SD 1.21%; 95% CI 7.33%-7.56%; low: 7.76%, SD 1.36%; 95% CI 7.52%-8.00%) (Bonferroni correction: high vs low, *P*=.08; mid vs low, *P*=.04; high vs mid, *P*>.99) (Figure 5B). Similar results were obtained among users with HbA_{1c} values of <8% in time bucket B_0 (high: 6.81%, SD 0.62%; 95% CI 6.75%-6.88%; mid: 6.86%, SD 0.57%; 95% CI 6.81%-6.90%; low: 6.79%, SD 0.61%; 95% CI 6.71%-6.86%) ($F_{2,1269}$=1.425; *P*=.24) and in time bucket B_4 (high: 6.72%, SD 0.74%; 95% CI 6.65%-6.80%; mid: 6.72%, SD 0.72%; 95% CI 6.67%-6.77%; low: 6.86%, SD 0.86%; 95% CI 6.75%-6.97%) (Bonferroni correction: high vs low, *P*=.08; mid vs low, *P*=.03; high vs mid, *P*>.99) (Figure 5C).

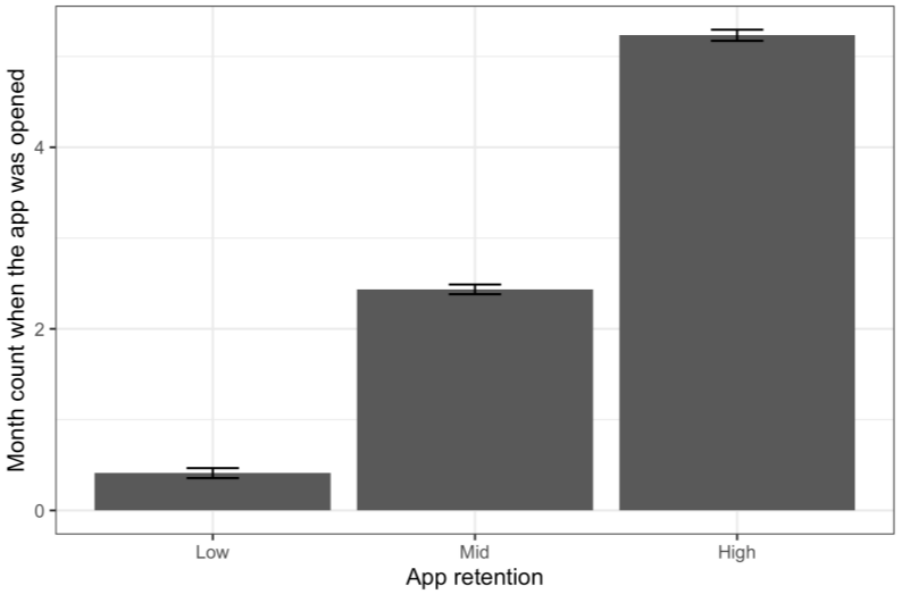
XSL•FO
RenderX

Figure 5. Users' mean HbA_{1c} in each time bucket. The error bars represent the standard errors among users in the group.



Users' retention behavior patterns remained the same in the second half of the tracking period (Figure 6). One-way ANOVA revealed significant differences in app retention across the 3 groups ($F_{2,1797}$=1082; P<.001), and subsequent post hoc analyses using Bonferroni correction revealed significant differences among the groups (high: 5.23 [SD 1.37] months, mid: 2.43 [SD 1.68] months, low: 0.41 [SD 0.97] months) (P<.001) for all comparison pairs.

Figure 6. Month count when the app was opened during the seventh to 12th month. The heights of the bars and the ranges of error bars represent the mean (SE) values in each group, respectively.



Discussion

Principal Findings

In this study, we found that patients with diabetes, who continuously used the Health2Sync app and were supported by HCPs who used the Health2Sync Patient Management Platform, presented a steeper reduction in HbA_{1c} levels and achieved lower HbA_{1c} levels after 1 year, compared to patients who dropped out and received traditional care for 1 year. Although we only used the app usage behavior data in the first 6 months to stratify the patients, their app usage patterns remained consistent in the following months. Additionally, our study shows positive results with regard to diabetes management from real-world data without prior RCT settings. While an RCT has an advantage over real-world evidence in providing homogeneous study groups for comparisons with designed inclusion and exclusion criteria, it is incapable of reflecting actual clinical practice where heterogeneous scenarios exist [22,23]. As digital interventions are to be applied to all patients,

XSL•FO
RenderX

we believe that our study with real-world evidence is more convincing in demonstrating efficacy.

In our study, both the mid and high retention groups presented a greater reduction in HbA_{1c} levels than their counterparts in the low retention group. In other words, only users who stopped using the app in the first month after adding a clinic or hospital as a partner saw poorer HbA_{1c} improvements than others. The beneficial effects of the Health2Sync app and Patient Management Platform might result from the in-app structured display and the general education or personalized feedback that further enhance users' health awareness, which were reported to be associated with better HbA_{1c} outcomes [13,24].

Although the majority of included subjects chose not to disclose their diabetes type, and the sample size was unbalanced across known diabetes types and retention groups, we performed a subgroup analysis to investigate the effect of app retention on different diabetes types, specifically type 1 and type 2. Based on changes in the raw HbA_{1c} values, we found that the reduction in the HbA_{1c} levels of only the patients with type 2 diabetes was related to app retention, while that of patients with type 1 diabetes seemed unrelated to app adherence (Multimedia Appendices 1 and 2 show the trend in app retention for patients with type 1 and type 2 diabetes, respectively). Many studies have focused on the efficacy of digital interventions for patients with type 2 diabetes or provided inadequate evidence of the efficacy of such interventions among patients with type 1 diabetes [7,9,14]. Some studies or analyses reported positive outcomes among patients with type 1 diabetics, but the effects were not as large as those among patients with type 2 diabetes [8,13,17].

Limitations

Our data also suggest such differences between the patients with type 1 and those with type 2 diabetes. However, considering the limited sample, future studies are needed to investigate whether the benefits from digital tools can also be applied to patients with type 1 diabetes, especially in real-world settings. Another study limitation is that we did not consider the differences in medication and daily behaviors such as exercise across patients, and these factors could have vital impacts on glycemic control. Future studies should also include these features in analyses.

With the growing prevalence of diabetes, it may become increasingly difficult to care for patients without increasing the number of HCPs. As such, digital solutions can play a critical role to enhance and scale diabetes care [25]. Moreover, during the COVID-19 pandemic, diabetes was found to be a common comorbidity, and patients with diabetes were found to be more vulnerable to COVID-19 [26-28]. A mobile app and platform that can enable HCPs to practice remote care should be helpful for glycemic control and can reduce the risk of people with diabetes being infected during visits to clinics or hospitals [29].

Conclusions

In conclusion, our retrospective analyses show that continuous usage of the Health2Sync app and Patient Management Platform was helpful for improving HbA_{1c} levels. Further studies are needed to reveal the efficacy of such interventions for specific diabetes types and to observe the effects beyond 1 year.

Authors' Contributions

YZT was responsible for the design of the analysis protocol, conducting statistical analyses, and drafting the manuscript. YTC, HYC, and KL assisted with the statistical analyses. All authors reviewed the final manuscript.

Conflicts of Interest

The Health2Sync mobile app and web-based Patient Management Platform are products of H2 Inc. YZT, YTC, and KL are full-time employees at H2 Inc, and KL supervises YZT and YTC. HYC received a consulting fee to assist with the analyses but otherwise declared no conflict of interest.

Multimedia Appendix 1
Diabetes type 1 users' mean HbA_{1c} in each time bucket. The error bars represent the standard errors among users in the group.
[PNG File , 129 KB - jmir_v23i7e23227_app1.png]

Multimedia Appendix 2
Diabetes type 2 users' mean HbA_{1c} in each time bucket. The error bars represent the standard errors among users in the group.
[PNG File , 137 KB - jmir_v23i7e23227_app2.png]

References

1. Global Report on Diabetes. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/204871 [accessed 2021-06-25]
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation

XSL•FO
RenderX

Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019 Nov;157:107843. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
3. Health Promotion Administration, Ministry of Health and Welfare. 國民營養健康狀況變遷調查2013-2016年成果報告. 2013. URL: https://www.hpa.gov.tw/Pages/ashx/File.ashx?FilePath=~/File/Attach/11145/File_12788.pdf [accessed 2021-06-25]
4. National Health Insurance Administration, Ministry of Health and Welfare. 2018年全民健康保險醫療費用前二十大疾病. 2019. URL: https://tinyurl.com/46yphwpw [accessed 2021-06-25]
5. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
6. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
7. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in Glycemic Control With Use of a Digital Therapeutic in Adults With Type 2 Diabetes: Cohort Study. JMIR Diabetes 2018 Feb 14;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
8. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
9. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
10. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of Mobile Apps to Support the Care of Patients With Diabetes Mellitus: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
11. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. CMAJ 2017 Mar 06;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]
12. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027 [FREE Full text] [doi: 10.1177/1932296817713506] [Medline: 28560898]
13. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
14. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of Digital Interventions for Improving Glycemic Control in Persons with Poorly Controlled Type 2 Diabetes: A Systematic Review, Meta-analysis, and Meta-regression Analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
15. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
16. Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What Do Adults with Type 2 Diabetes Want from the "Perfect" App? Results from the Second Diabetes MILES: Australia (MILES-2) Study. Diabetes Technol Ther 2019 Jul;21(7):393-399. [doi: 10.1089/dia.2019.0086] [Medline: 31166804]
17. Osborn CY, van Ginkel JR, Rodbard D, Heyman M, Marrero DG, Huddleston B, et al. One Drop | Mobile: An Evaluation of Hemoglobin A1c Improvement Linked to App Engagement. JMIR Diabetes 2017 Aug 24;2(2):e21 [FREE Full text] [doi: 10.2196/diabetes.8039] [Medline: 30291059]
18. Nathan DM, Turgeon H, Regan S. Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia 2007 Nov;50(11):2239-2244. [doi: 10.1007/s00125-007-0803-0] [Medline: 17851648]
19. R: a language and environment for statistical computing. Global Biodiversity Information Facility. URL: https://www.R-project.org/ [accessed 2021-06-25]
20. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
21. Kuznetsova A, Brockhoff PB, Christensen RHB. Package: Tests in Linear Mixed Effects Models. J Stat Soft 2017;82(13). [doi: 10.18637/jss.v082.i13]
22. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value Health 2017;20(7):858-865 [FREE Full text] [doi: 10.1016/j.jval.2017.03.008] [Medline: 28712614]

XSL•FO
RenderX

23. Kim H, Lee S, Kim JH. Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. J Korean Med Sci 2018 Aug 20;33(34):e213 [FREE Full text] [doi: 10.3346/jkms.2018.33.e213] [Medline: 30127705]
24. Tang YH, Pang SM, Chan MF, Yeung GS, Yeung VT. Health literacy, complication awareness, and diabetic control in patients with type 2 diabetes mellitus. J Adv Nurs 2008 Apr;62(1):74-83. [doi: 10.1111/j.1365-2648.2007.04526.x] [Medline: 18352966]
25. Ramchandani N. Virtual Coaching to Enhance Diabetes Care. Diabetes Technol Ther 2019 Jun;21(S2):S248-S251. [doi: 10.1089/dia.2019.0016] [Medline: 31169430]
26. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020 Apr;8(4):e21 [FREE Full text] [doi: 10.1016/S2213-2600(20)30116-8] [Medline: 32171062]
27. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
28. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
29. Gupta R, Ghosh A, Singh AK, Misra A. Clinical considerations for patients with diabetes in times of COVID-19 epidemic. Diabetes Metab Syndr 2020;14(3):211-212 [FREE Full text] [doi: 10.1016/j.dsx.2020.03.002] [Medline: 32172175]

Abbreviations

ANOVA: analysis of variance
HbA_{1c}: hemoglobin A_{1c}
HCP: health care professional
RCT: randomized controlled trial

Edited by R Kukafka; submitted 05.08.20; peer-reviewed by B Carter, G Darville; comments to author 27.08.20; revised version received 22.10.20; accepted 24.05.21; published 15.07.21.

Please cite as:
Tu YZ, Chang YT, Chiou HY, Lai K
The Effects of Continuous Usage of a Diabetes Management App on Glycemic Control in Real-world Clinical Practice: Retrospective Analysis
J Med Internet Res 2021;23(7):e23227
URL: https://www.jmir.org/2021/7/e23227
doi:10.2196/23227
PMID:34264192

©Yu-Zhen Tu, Ya-Ting Chang, Hung-Yi Chiou, Ken Lai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 15.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Uptake of and Engagement With an Online Sexual Health Intervention (HOPE eIntervention) Among African American Young Adults: Mixed Methods Study

Alicia Williamson[1], BA; Andrea Barbarin[2], MS, JD, PhD; Bettina Campbell[3], LMSW; Terrance Campbell[3,4], MA Ed, MSIM; Susan Franzen[5], MS; Thomas M Reischl[5], PhD; Marc Zimmerman[6,7], PhD; Tiffany Christine Veinot[1,6], MLS, PhD

[1]School of Information, University of Michigan, Ann Arbor, MI, United States

[2]IBM Watson Health, Ann Arbor, MI, United States

[3]YOUR Center, Flint, MI, United States

[4]TigerLIFE, University of Memphis, Memphis, TN, United States

[5]Prevention Research Center of Michigan, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[6]Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[7]Department of Psychology, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Tiffany Christine Veinot, MLS, PhD
School of Information
University of Michigan
4314 North Quad
105 S State Street
Ann Arbor, MI, 48109-1285
United States
Phone: 1 734 615 8281
Fax: 1 734 615 8281
Email: tveinot@umich.edu

Abstract

Background: Regarding health technologies, African American young adults have low rates of uptake, ongoing usage, and engagement, which may widen sexual health inequalities.

Objective: We aimed to examine rates of uptake and ongoing usage, and factors influencing uptake, ongoing usage, and engagement for a consumer health informatics (CHI) intervention for HIV/sexually transmitted infection (STI) prevention among African American young adults, using the diffusion of innovation theory, trust-centered design framework, and O'Brien and Toms' model of engagement.

Methods: This community-based participatory mixed methods study included surveys at four time points (n=315; 280 African American participants) among young adults aged 18 to 24 years involved in a blended offline/online HIV/STI prevention intervention (HIV Outreach, Prevention, and Education [HOPE] eIntervention), which was described as a "HOPE party." Qualitative interviews were conducted with a subset of participants (n=19) after initial surveys and website server logs indicated low uptake and ongoing usage. A generalized linear mixed-effects model identified predictors of eIntervention uptake, server logs were summarized to describe use over time, and interview transcripts were coded and thematically analyzed to identify factors affecting uptake and engagement.

Results: Participants' initial self-reported eIntervention uptake was low, but increased significantly over time, although uptake never reached expectations. The most frequent activity was visiting the website. Demographic factors and HOPE party social network characteristics were not significantly correlated with uptake, although participant education and party network gender homophily approached significance. According to interviews, one factor driving uptake was the desire to share HIV/STI prevention information with others. Survey and interview results showed that technology access, perceived time, and institutional and technological trust were necessary conditions for uptake. Interviews revealed that factors undermining uptake were insufficient promotion and awareness building, and the platform of the intervention, with social media being less appealing due to previous

XSL•FO
RenderX

negative experiences concerning discussion of sexuality on social media. During the interaction with the eIntervention, interview data showed that factors driving initial engagement were audience-targeted website esthetics and appealing visuals. Ongoing usage was impeded by insufficiently frequent updates. Similarly, lack of novelty drove disengagement, although a social media contest for sharing intervention content resulted in some re-engagement.

Conclusions: To encourage uptake, CHI interventions for African American young adults can better leverage users' desires to share information about HIV/STI prevention with others. Ensuring implementation through trusted organizations is also important, though vigorous promotion is needed. Visual appeal and targeted content foster engagement at first, but ongoing usage may require continual content changes. A thorough analysis of CHI intervention use can inform the development of future interventions to promote uptake and engagement. To guide future analyses, we present an expanded uptake and engagement model for CHI interventions targeting African American young adults based on our empirical results.

(J Med Internet Res 2021;23(7):e22203) doi:10.2196/22203

KEYWORDS

HIV prevention; consumer health informatics; sexual health; health equity; technology adoption; technology usage

Introduction

Background

HIV and other sexually transmitted infections (STIs) are health concerns for African American young adults. In 2018, African Americans constituted 13% of the population but 42% of new HIV diagnoses in the United States [1], with young adults making up about 21% of new HIV diagnoses [2]. African American young adults are vulnerable to other STIs, having the highest rates of gonorrhea of all racial groups [3]. Consumer health informatics (CHI) technologies can support HIV/STI prevention by targeting outcomes such as sexual health literacy and HIV/STI testing rates [4].

However, previous experience with patient portals shows that CHI interventions may not achieve uptake and engagement among African Americans [5,6]. In an HIV/STI context, this may widen sexual health inequalities [7,8]. Numerous CHI interventions exist to prevent HIV [9-15] and STIs [16,17] among African American youth. Additionally, digital CHI interventions targeting African American populations can be effective at promoting safer sex behaviors [18,19] and can increase access to sexual health information [20,21]. However, few identify factors driving uptake and engagement [22]. While interventions have been effective at targeting African American youth [9,13,15], limited uptake and engagement can impede the promotion of behaviors, such as condom use [23,24], or the development of sexual health knowledge [25,26]. We define *uptake* as initiating contact with a CHI intervention, which is differentiated from *ongoing usage* of that intervention [27]. Uptake is rarely investigated, inconsistently defined, and often underreported [28,29]. *Engagement* has been defined by some as interactions over time [30], although information scientists use the term to demarcate the period in which a user interacts with an intervention [31]. Given our interest in experiences of CHI interventions, we use the latter definition of engagement in this study.

Factors driving uptake may explain disparities between racial groups in CHI intervention use [32]. Studies that examine uptake among African Americans suggest that it may be impeded by recruitment barriers, power differentials between researcher/clinician and participants, or researchers' poor cultural competence [33]. Medical distrust stemming from discrimination and historical mistreatment also plays a role [34]. However, targeted recruitment and individual tailoring may improve uptake among African American young adults. Peer-driven recruitment [35,36] and partnering with community-based organizations [37] may increase uptake. Tailoring web-based interventions to users' needs, interests, personalities, and contexts [38], and tailored messages from a personalized source [39] can also improve uptake.

CHI interventions should also foster ongoing usage. It can be difficult to differentiate between factors motivating uptake and ongoing usage; few studies have examined this. Some evidence shows that African American young adults' engagement may be impeded by distrust, lack of time, limited technology access, and limited cultural relevance [32].

Some CHI interventions have successfully engaged African American users via social media, peer-led interventions, and designing to capture user attention. Social media-based HIV-prevention interventions have shown low attrition [14]. Peer-led CHI interventions have maintained engagement through existing social networks [14]. High levels of interest [40] and perceived usefulness may also drive ongoing usage [41].

Yet, there is limited understanding of what motivates uptake and engagement in African American young adults participating in CHI interventions to prevent HIV/STIs. We identify motivating factors for a tailored website and social media accounts promoting safer sex for African American young adults, framing our analyses with the theories described below.

Theoretical Frameworks

Diffusion of Innovations

The diffusion of innovations (DOI) theory describes a process of communicating about innovations among members of social systems [42]. Communication can lead to adoption (deciding to use an intervention; this aligns with our definition of uptake). Personal characteristics, such as demographics, influence adoption. For example, people with more education and higher social participation tend to adopt innovations earlier. Furthermore, homophily (social similarity) with adopters increases the likelihood that nonadopters will make the

XSL•FO
RenderX

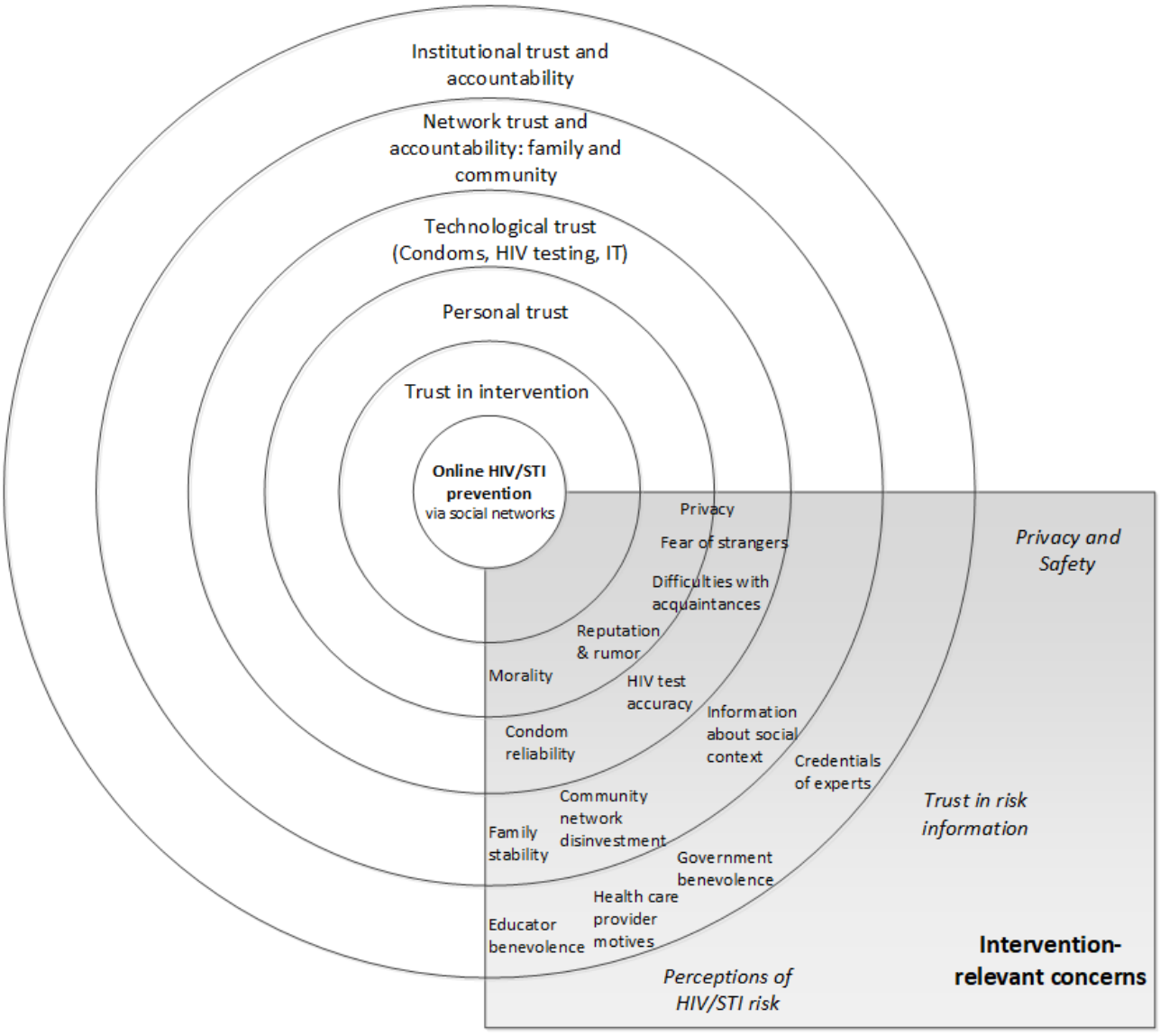
subsequent decision to adopt an innovation. We used DOI to identify variables that may predict uptake.

Several HIV-prevention interventions have used the DOI theory [43]. Similarly, it informed the design of the HIV Outreach, Prevention, and Education (HOPE) eIntervention in this study. This intervention asked African American young adults to invite peers to HOPE parties (described below), which introduced the innovations (both the eIntervention and healthy sexual behaviors).

Trust-Centered Design Framework

The trust-centered design framework (Figure 1) [12] outlines trust-based needs in relation to CHI interventions at the personal, group, technological, and institutional levels for African American young adults. It centers trust in CHI interventions, calls for addressing trust-related requirements at multiple user-experience levels, and highlights forms of trust that may influence uptake and engagement. We used this framework to inform HOPE technology design and implementation strategies, and to analyze qualitative data in this study.

Figure 1. Trust-centered design framework. IT: information technology; STI: sexually transmitted infection.

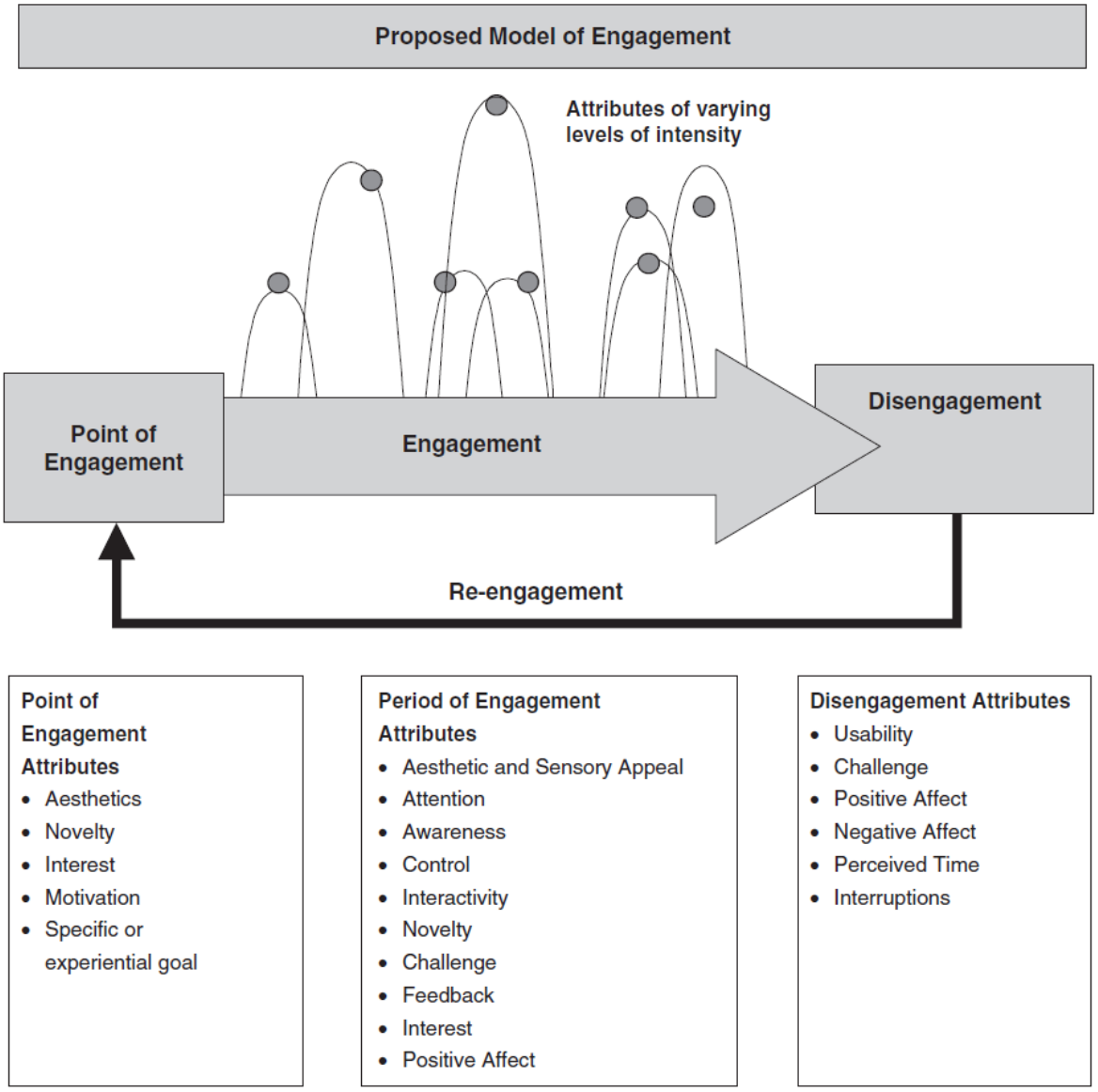


User Engagement Model

O'Brien and Toms' model (Figure 2) focuses specifically on the period, or session, of interacting with an intervention. Focusing on websites as an intervention type, it differentiates the following four stages of engagement: point of engagement, period of engagement, disengagement, and re-engagement [31]. It posits that there are different drivers at each stage, grouped into (1) website features and (2) user experience features. Points of engagement and re-engagement are influenced by web features of esthetics and novelty, and user features including motivation. The period of engagement is influenced by sensory appeal, interactivity, novelty, challenge, feedback, attention, awareness, control, and positive affect. Disengagement is driven by website features of usability and challenge, and user features of positive/negative affect, perceived time, and interruptions. These factors vary in influence over time. We used this model to analyze qualitative interview data.

XSL•FO
RenderX

Figure 2. O'Brien and Toms' model of engagement [31]. Reprinted with permission.



Study Objectives

The aim of this study was to analyze uptake, ongoing usage, and engagement with a CHI intervention, the *HOPE eIntervention*, intended for African American young adults, which included a website and social media components. The research questions were as follows: RQ1: How much did HOPE participants use the eIntervention (uptake and ongoing usage)? RQ2: What factors were related to eIntervention uptake (initiating contact with the intervention)? RQ3: What factors were associated with ongoing usage and engagement while a participant interacted with the eIntervention?

Methods

Intervention Context

This mixed-methods, community-based participatory research (CBPR) [44] study included surveys and interviews with African American young adults who attended a community-based HIV/STI prevention intervention called a *HOPE party.* Regarding the CBPR approach, YOUR Center, a local faith-based nonprofit organization, was involved in study design, recruitment, implementation, data collection, and result dissemination, and the Genesee County Health Department assisted in recruitment; this is consistent with use of the CBPR approach in health informatics [44]. The team had monthly meetings during the intervention timeframe, and community members were involved in the design of the HOPE eIntervention, from its inception to its implementation. Furthermore, YOUR Center staff led the in-person interventions, and community members assisted by leveraging their social networks for recruitment. A community advisory board helped to integrate community needs throughout the project; this also aligns with CBPR practice [44]. Further information about the partnership can be found in a previous report [44].

Intervention Description

HOPE parties were one-time, interactive, face-to-face HIV/STI prevention sessions hosted by African American young adults in their homes for their social networks. Attendees listened to

XSL•FO
RenderX

didactic material, engaged in educational activities, and discussed safe sex practices with demonstrations of safe sex tools [41]. Local resources were also discussed, including information about the eIntervention. YOUR Center staff were trained to conduct the parties and used a facilitator's manual to ensure that core content was covered in each party; fidelity assessments were also used at each party. However, as expected, there was some variation across parties. Recruitment was conducted at Genesee County STI clinics by trained UM School of Public Health, Health Professions and Studies students and community residents. YOUR Center staff also recruited participants through naturally occurring face-to-face professional and personal social networks, and snowball recruitment in which hosts recruited others.

The eIntervention, available to HOPE party participants, aimed to reinforce information presented at the parties and foster commitment to HIV/STI prevention. Together, the face-to-face parties and eIntervention comprised a blended offline/online intervention. The eIntervention was designed in collaboration with YOUR Center's staff and community advisory board, and informed by design-oriented focus groups [12]. Based on the trust concerns when communicating about HIV/STIs that were raised in the trust-centered design framework developed as part of this study [12], the website allowed participants to remain anonymous to foster interactions that preserved privacy. Indeed, community partners expressed concerns about the acceptability of user accounts based on their interaction with the target audience for the intervention; thus, profiles were not required in order to access website content. Due to issues raised around inconsistent access to trusted information about sexual health [12,45], partnering with YOUR Center was intended to create an environment where African American young adult users would trust the information based on its institutional affiliation. Additionally, the blog features of the website allowed for users to access credible information via these short and casual posts. The website had a theme for mobile devices such as cellphones. A website plug-in detected what device was used to access the intervention and loaded either a desktop or mobile theme based on this detection. Users could toggle between the mobile and desktop versions of the website if desired.

At each HOPE party, YOUR Center staff and/or University of Michigan research staff introduced the eIntervention and distributed handouts that contained the website URL and directions on how to access social media accounts. The eIntervention consisted of a website with a blog, a Twitter account, and a Facebook page (examples of posts are included in Multimedia Appendix 1, Multimedia Appendix 2, and Multimedia Appendix 3). The website was updated at least once monthly. Facebook content included status updates, external links, photos, sexual health information, and/or questions. Tweets offered safer sex advice and memes reinforcing safer sex norms. Three social media contests incentivized eIntervention information sharing via social networks.

Data Collection

HOPE party participants (n=315; 280 African American; some participants invited non-African American friends to parties) completed four surveys during the study. These included (1) at the party, in-person (baseline); (2) 3 months postparty, online/via phone (T1); (3) 6 months postparty, online/via phone (T2); and (4) 12 months postparty, online/via phone (T3). There were 57 HOPE parties over 3 years, with 1 to 12 participants at each who responded to surveys. Follow-up survey requests were first attempted over the phone, and those participants whom we could not reach were contacted via US mail and by email. We attempted to contact all survey participants regardless of the status of completion for the other follow-up waves. Participants received US $10 for completing the baseline, T1, and T2 questionnaires, and US $15 for completing the T3 questionnaire. Participants who completed all four surveys were entered into a draw for a US $100 gift card. The survey asked about demographic characteristics, social network information, and questions regarding use of the HOPE eIntervention. Surveys were developed to measure network assessments such as network composition (gender homophily) [46], social influence of networks based on density [47], and type of social tie (ie, weak or strong) [48] as predictors of HOPE website use. Survey questions used in the data analysis are detailed further in Multimedia Appendix 4.

After survey and website server logs indicated low uptake and engagement, individual in-depth semistructured interviews [49] (n=19) were conducted with African American young adult party participants. Interviews were conducted at least 1 year after each participant's HOPE party. One graduate student (AB) from the University of Michigan who was trained in qualitative interviewing conducted the interviews using a semistructured interview guide that accommodated personalized follow-up questions. Using theoretical sampling [50], participants reporting low, medium, and high levels of eIntervention use on the survey were recruited via telephone or email. We contacted approximately 131 participants in total (56 participants with low levels of eIntervention use, 62 participants with medium levels of eIntervention use, and 12 participants with high levels of eIntervention use). Of the 131 participants, 19 participated in interviews (15% response rate).

We concluded interviews when we reached data saturation such that new data were no longer contributing new empirical insights [51]. Others have shown that data saturation can be achieved with as few as six to eight people in a relatively homogenous group such as that in this study (African American young adults in one geographic area) [52]. A timeline of the partnership with YOUR Center, development of the intervention, HOPE parties, follow-up surveys, and interviews is included in Multimedia Appendix 5.

Google Analytics measures included numbers of sessions, users, and page views; session duration; and trends. HOPE Twitter and Facebook data included numbers of followers/likes.

Measures

Dependent Variable

At each survey wave, the main outcome in the individual-level statistical model for predicting uptake was a binary variable for having used the HOPE website or social media since the party. For the party-level model, this outcome was an average of the

XSL•FO
RenderX

number of party attendees who used the eIntervention at each wave.

Independent Variables-Individual Level

Education

The DOI theory states that more education increases the speed of using innovations [42]; thus, educational level was a predictor. In the model, the variable had the following two levels: (1) less than high school diploma or General Education Diploma (GED) and (2) high school diploma or GED, some college, associate's degree, or bachelor's degree.

Gender

Gender was included as a binary predictor (male/female). This was included since men are more likely to have favorable technology attitudes [53], though these attitudes about technologies vary by age [54]. Further, women use health websites more [55] and are more likely to search for health information online [56].

Time Point

Participants' use by time point was included because we observed increases in overall usage of the eIntervention over time.

Independent Variables-Party Level

Gender Homophily

The DOI theory predicts that homophily increases the odds of adoption, and our previous work has shown that gender homophily is associated with HIV testing behavior [46,57]. Thus, gender homophily was used as an independent variable, which was calculated using the Krackhardt E-I index network measure [58] as the number of females minus the number of males at the party divided by total participants [58].

Network Density

Network density [47] was included since it is linked to social influence [57]. This standard network measure [59] quantifies the number of participants at a party who have relationships divided by the number of total possible relationships between party members. Higher levels of network density are associated with health behavior transmission between youth [60,61]. This was calculated as the total number of attendee ties (any self-reported relationship before the party) divided by total possible ties [62].

Strong Tie Proportion

Strong ties [48] are also linked to social influence. They tell us the proportion of relationships between people at a party that are based on close interpersonal relationships [57]. Using a standard measure of tie strength, this survey question asked participants how close they felt to other party attendees, with the options being very close (strong ties), somewhat close, and not close [63]. There is ample evidence that people with whom one is emotionally intimate, such as close friends, influence one's sexual risk behaviors [64-66]. This was measured by the number to whom each attendee felt very close divided by the number of existing ties between party members.

Qualitative Interview Protocol

Interviews included (1) a demographic and technology use questionnaire, and (2) interviews about party experience, eIntervention uptake and engagement, internet and social media use, and HIV/STI-related communication practices.

Data Analysis

Survey Analysis

Survey data were explored using descriptive statistics. Data analysis was conducted after the conclusion of T3. To identify uptake predictors, Stata was used to fit a generalized linear mixed-effects model [67] with binomial distribution to account for (1) the binary outcome variable (eIntervention use) and (2) repeated measures across the three time points. This method accounts for both within-participant and across-participant variability. The coefficients in this model can be interpreted as the log odds of predicted HOPE eIntervention use by variable. Due to our focus on African Americans, only the sample of African Americans (n=280) was entered into this model.

A second generalized linear mixed-effects model with a beta distribution using party-level indicators was fit. Uptake trends were also analyzed with respect to independent party-level variables such as gender homophily, network density, and strong tie proportion. There was a random effect for the person-level variable and were fixed effects for party ID (identifier) and within-party variables (gender homophily, strong tie proportion, and cohesion density) to predict average uptake at the party level while controlling for party size. For this model, the entire sample (n=315) was entered into the model as parties had both African American and non-African American attendees, although African Americans were the focus of the intervention design.

Interview Data Analysis

Interviews, averaging 43 minutes, were audio-recorded and transcribed. Among the 19 interview participants, 8 (42%) had low self-reported usage levels, 6 (32%) had medium self-reported usage levels, and 5 (26%) had high self-reported usage levels, with representation across usage levels. Transcripts were analyzed using deductive and inductive approaches in NVivo. Deductively, O'Brien and Toms' [31] model of engagement and the trust-centered design framework [12] provided codes; line-by-line coding resulted in inductive codes. One coder initially coded the interviews, and a second coder reviewed the codes and reached consensus with the first. This resulted in emergent categories regarding information sharing, party experience, eIntervention experience, and social media [68]. Analytical memos were used to develop themes after reviewing and combining or collapsing codes [68]. Categories were also developed deductively using O'Brien and Toms' model of engagement in the period, or session, in which users interacted with an intervention. Their model includes initial engagement, period of engagement, re-engagement, and disengagement. In this context, initial engagement is conceptualized as the beginning of the interaction with the eIntervention; period of engagement is the remaining time through which a participant uses the eIntervention; re-engagement is when a participant returns to engage with the

XSL•FO
RenderX

eIntervention after days, weeks, or months; and disengagement is when a participant discontinues use of the eIntervention in a given session. Re-engagement and disengagement can be thought of as components of ongoing usage.

Website/Social Media Usage Data Analysis

Frequency counts were compiled for Google Analytics, Twitter, and Facebook data.

Results

Characteristics of the Participants

As Table 1 shows, 280 African American participants responded to the survey, and 19 African American young adult interviewees were drawn from this pool of respondents. Of the 280 survey respondents, 164 (58.6%) were female and 110 (39.3%) completed high school or equivalent, and the mean age was 21.13 years. Of the 19 interview participants, 14 (74%) were female, with similar education levels and a mean age of 24 years. As Table 2 shows, there were 57 parties in total, with an average of 5.63 (SD 2.56) participants per party who responded to the survey. Of the total 315 attendees who responded to the survey, 280 (88.9%) were African American.

Table 1. Summary of descriptive statistics for the African American sample at baseline.

Characteristic[a]	Quantitative survey (n=280)	Qualitative interview (n=19)
Gender, n (%)		
Male	112 (40%)	5 (26%)
Female	164 (59%)	14 (74%)
Missing	4 (1%)	0 (0%)
Age at the time of the HOPE[b] party, mean (SD)	21.13 (2.25)	24 (3.317)
Education level at the time of the HOPE party, n (%)		
Less than high school diploma or GED[c]	80 (29%)	1 (5%)
High school diploma or GED	110 (39%)	8 (42%)
Some college, associate's degree, or bachelor's degree	85 (30%)	10 (53%)
Missing	5 (2%)	0 (0%)
Host, n (%)		
Yes	25 (9%)	3 (16%)
No	255 (91%)	16 (84%)
Missing	0 (0%)	0 (0%)
Employment status[d], n (%)		
Full-time work	30 (11%)	4 (21%)
Part-time work	42 (15%)	1 (5%)
Full-time student	34 (12%)	3 (16%)
Part-time student	12 (4%)	2 (11%)
Unemployed	165 (59%)	10 (53%)
Other	14 (5%)	2 (11%)
Missing	2 (0%)	0 (0%)

[a]All survey respondents identified as African American. Of the 19 qualitative interview respondents, 18 (95%) identified as African American and 1 (5%) identified as multiracial.

[b]HOPE: HIV Outreach, Prevention, and Education.

[c]GED: General Education Diploma.

[d]Participants could choose multiple employment statuses.

XSL•FO
RenderX

Table 2. Summary of descriptive statistics for baseline party characteristics.

Variable	Value
Total parties[a], n (%)	57 (100)[b]
Total attendees per party[c], mean (SD; min, max)	5.63 (2.56; 1, 12)
Average number of African American young adults per party[c], mean (SD; min, max)	5 (2.16; 1, 11)
Average number of non-African American young adults per party[c], mean (SD; min, max)	0.57 (1.22; 0, 6)
Average number missing per party[c], mean (SD; min, max)	0.05 (0.23; 0, 1)
Average percentage of African American young adults per party[c], mean (SD)	91.8% (16.7%)
Average percentage of non-African American young adults per party[c], mean (SD)	7.4% (16.0%)
Average percentage missing per party[c], mean (SD)	0.9% (3.8%)
Total attendees[c], n (%)	315 (100%)
Total African American young adult attendees[c], n (%)	280 (88.9%)
Total non-African American young adult attendees[c], n (%)	32 (10.1%)
Total missing demographics[c], n (%)	3 (1.0%)
Network density[c], mean (SD; min, max)	0.80 (0.19; 0.4, 1.0)
Strong tie proportion[c], mean (SD; min, max)	0.27 (0.17; 0.0, 0.8)
Party gender homophily[c], mean (SD; min, max)	0.23 (0.58; −1.0, 1.0)

[a]Includes party participants who did not complete a survey.

[b]There were 57 total parties, but one party did not have any respondents fill out the survey at baseline, so 56 were included in the analyses.

[c]Includes only party participants who completed a survey.

RQ1 Results

Survey Results

Initial uptake of the eIntervention and ongoing usage were low (Table 3). Self-reported use increased at subsequent time points. Initially, only 21 of 280 (8%) participants reported ever using the eIntervention (uptake), with an increase to 35 of 280 participants (13%) at T2 and 42 of 280 participants (15%) at T3. The most frequent engagements were visiting the website (T1: n=9, T2: n=24, T3: n=34), visiting the Facebook page (T1: n=1, T2: n=8, T3: n=11), and tweeting (T1: n=2, T2: n=4, T3: n=5). However, large amounts of survey data were missing either due to nonresponse at that time point or due to skipping that question (Table 3). Of the 359 total HOPE party participants, 315 participants took at least one survey. Regarding response at each time period, among 359 HOPE party participants, 178 responded at T1 (49.6% response rate), 180 responded at T2 (50.1% response rate), and 186 responded at T3 (51.8% response rate). Among the 280 African American party participants who ever answered the survey, 20 (7%) answered at only T1, 14 (5%) answered at only T2, and 16 (6%) answered at only T3. Table 3 outlines responses from the 280 African American respondents who ever responded to a survey.

XSL•FO
RenderX

Table 3. HIV Outreach, Prevention, and Education (HOPE) eIntervention use (N=280).

Variable	Time point 1, n (%)	Time point 2, n (%)	Time point 3, n (%)
Use of the HOPE[a] eIntervention since attending the HOPE party			
Yes	21 (7.5%)	35 (12.5%)	42 (15.0%)
No	140 (50.0%)	127 (45.4%)	123 (43.9%)
Missing[b]	119 (42.5%)	118 (42.1%)	115 (41.1%)
What online HOPE activities did you do?[c]			
Visiting or using the HOPE website[d]	16 (76.2%)	24 (68.6%)	34 (80.1%)
Visit the Facebook page	1 (4.8%)	8 (22.9%)	11 (26.2%)
Post or blog on the HOPE website	0 (0%)	0 (0%)	1 (2.4%)
Tweet on Twitter	2 (9.5%)	4 (11.4%)	5 (12.0%)
Post on Facebook	0 (0%)	3 (8.6%)	1 (2.4%)
Other	0 (0%)	3 (8.6%)	1 (2.4%)
Why didn't you use the HOPE eIntervention in the last 30 days?[c]			
No time	54 (41.9%)	60 (47.2%)	58 (46.0%)
I did not have computer access	38 (29.5%)	31 (24.4%)	36 (28.6%)
I did not want to	13 (10.1%)	23 (18.1%)	14 (11.1%)
Other	24 (18.6%)	18 (14.2%)	17 (13.5%)
I did not know about it[e]	10 (7.8%)	10 (7.9%)	7 (5.6%)
I forgot about it[e]	8 (6.2%)	5 (3.9%)	6 (4.8%)

[a]HOPE: HIV Outreach, Prevention, and Education.

[b]Missing responses were either due to nonresponse at that time point or due to skipping the question.

[c]Participants were able to choose multiple responses.

[d]In the first time point (T1), participants were asked for their general use of the HOPE website as well as if they visited the website. In T2 and T3 surveys, the "general use" question was eliminated due to overlap.

[e]Open responses for "other" were categorized by "I did not know about it" and "I forgot about it."

Log File and Social Media Results

According to the Google Analytics report, the website had 2432 sessions and 5754 total page views from July 2011 to March 2014, with an average session duration of 2 minutes 52 seconds. The Facebook page was updated with five to seven posts weekly and had 81 followers. HOPE Twitter was active from July 2011 to April 2014, with 131 followers. Inconsistency in reported usage may be due to missing survey responses or to eIntervention use by non-HOPE party participants.

RQ2 Results

Survey Results

To identify uptake predictors across the three time points, we fitted a generalized linear mixed-effects model at the individual and party levels (Table 4). Individual-level results suggested that having some college education was a marginally significant predictor of eIntervention use. No other individual-level factors predicted uptake in our analysis of HOPE eIntervention use after the conclusion of T3 surveys, although there was an increase in usage from T1 to T2, as well as an increase from T1 to T3.

XSL•FO
RenderX

Table 4. Generalized linear mixed-effects model results.

Model	Coefficient	SE	Z	*P* > |z|	Confidence interval	
					Lower	Upper
Individual-level model among African American party participants (n=280)						
Gender (female)	0.33	0.45	0.73	.47	−0.56	1.22
Time point 1 to 2	0.89	0.37	2.40	.01	0.16	1.61
Time point 2 to 3	1.12	0.37	3.05	.002	0.40	1.84
High school diploma or GED[a]	0.59	0.48	1.22	.22	−0.36	1.54
Some college, associate's degree, or bachelor's degree	0.93	0.48	1.92	.05	−0.02	1.88
Network density	0.51	1.01	0.51	.61	−1.46	2.49
Strong tie proportion	−0.53	1.25	−0.42	.67	−2.98	1.93
Party gender homophily	0.995	0.72	1.38	.17	−0.42	2.41
Party-level model among all party participants (n=315)						
Network density	−0.82	0.90	−0.92	.36	−2.58	0.93
Strong tie proportion	1.46	0.98	1.49	.14	−0.46	3.37
Gender homophily	0.05	0.03	1.81	.07	−0.004	0.11

[a]GED: General Education Diploma.

In the party-level model, gender homophily was the only marginally significant group-level variable, indicating a weak positive effect of a high proportion of females (gender homophily) on average uptake rates across participants in a given party.

As shown in Table 3, most participants who reported not having used the HOPE eIntervention in the past 30 days (of whom 100% had never used the intervention) at T1-T3 said this was due to lack of time (54-58 participants, 41.9%-47.2%). Following this, in frequency, were not having access to a computer (31-38 participants, 24.4%-29.5%), not wanting to (13-23 participants, 10.1%-18.1%), and other (17-24 participants, 13.5%-18.6%). Other responses concerned forgetting about the eIntervention or not being aware of it. Given their frequency, we interpreted perceived time and technology access as "necessary conditions" for uptake. Necessary conditions are facts that must be true for intervention uptake; however, they are insufficient on their own to promote uptake.

Interview Results

Individual Characteristics

Awareness

At least three interview participants were unaware of the eIntervention, although one mentioned visiting the website to see what was on it (Multimedia Appendix 6). Three participants suggested that uptake could be improved through more promotion.

Motivation

Some used the eIntervention with the goal of referring members of their community or family to it, believing they would benefit from learning about prevention (Multimedia Appendix 6). Furthermore, 12 interviewees shared information learned from the website with friends and family. Conversations with friends regarding HOPE usually involved sharing information about HIV/STI symptoms or testing locations (Multimedia Appendix 6).

Necessary Conditions

Necessary conditions were not typically enough to motivate uptake of the eIntervention, but their lack meant that people would not use it.

Technology Access

According to survey results, while some interview participants had internet access on varied devices, three did not have high-quality technology access, with two accessing it through others. Some participants also primarily accessed the internet on a mobile device; this limited their activities (Multimedia Appendix 6).

Perceived Time

Like survey respondents, some interviewees said they were too busy from work, school, and/or child care for the internet in general or for the HOPE eIntervention in particular (Multimedia Appendix 6).

Trust-Institutional

YOUR Center, which facilitated the parties, was considered a trustworthy information source (Multimedia Appendix 6). This perception was aided by personal connections. Three participants mentioned knowing someone linked to YOUR Center, and participants mentioned connections between HOPE parties and their churches. Attending parties organized and facilitated by individuals from trusted institutions made participants more likely to see the information from both the parties and eIntervention as credible.

XSL•FO
RenderX

Trust-Technological

Participants had trust in the website's information, a form of technological trust [12,15]. Seven participants mentioned understanding the information easily, perceiving it as clear and having an appropriate length (Multimedia Appendix 6). However, participants mentioned negative experiences on social media that may have shaped their perceptions of discussing sexual health information online, potentially resulting in lower uptake of eIntervention social media components. People expressed concern about potential technology-facilitated privacy breaches and online fighting. Furthermore, gossip was a major concern. Ten discussed negative experiences with "Facebook Exposed" pages, a user-generated network of pages featuring incendiary posts disclosing personal information about sexual promiscuity, sexual identity, and HIV/STI status.

RQ3 Results

Interview Results

Initial Engagement

Esthetics and Sensory Appeal

Participants responded positively to the website's esthetics. Four liked its visual layout, stating it was not too simple and had multiple colors (Multimedia Appendix 6).

Challenge/Ease of Use

Participants valued that information on the website was direct, short, and engaging. Its content was easy to use. Compared to other websites about sexual health, the HOPE website was easier to use and thus more accessible (Multimedia Appendix 6).

Period of Engagement

Interactivity

The interactivity present on the website, such as the blog, was appreciated. However, three participants suggested making the website more interactive by imitating social media platforms. One suggested user profiles, much like social networking sites (Multimedia Appendix 6). This approach could also offer *novelty.*

Disengagement

Novelty

As Multimedia Appendix 6 shows, two participants who were initially regular website users stopped visiting due to a lack of consistent updates.

Re-engagement

Promotion of the Intervention Through Contests

Social media contests were intended to incentivize participants to visit the website to share information with peers; these fostered some limited re-engagement. Two interviewees participated in the contest, and one male participant said that he was interested in the financial reward and thought it was a good way to engage people with the content (Multimedia Appendix 6).

Discussion

Principal Results

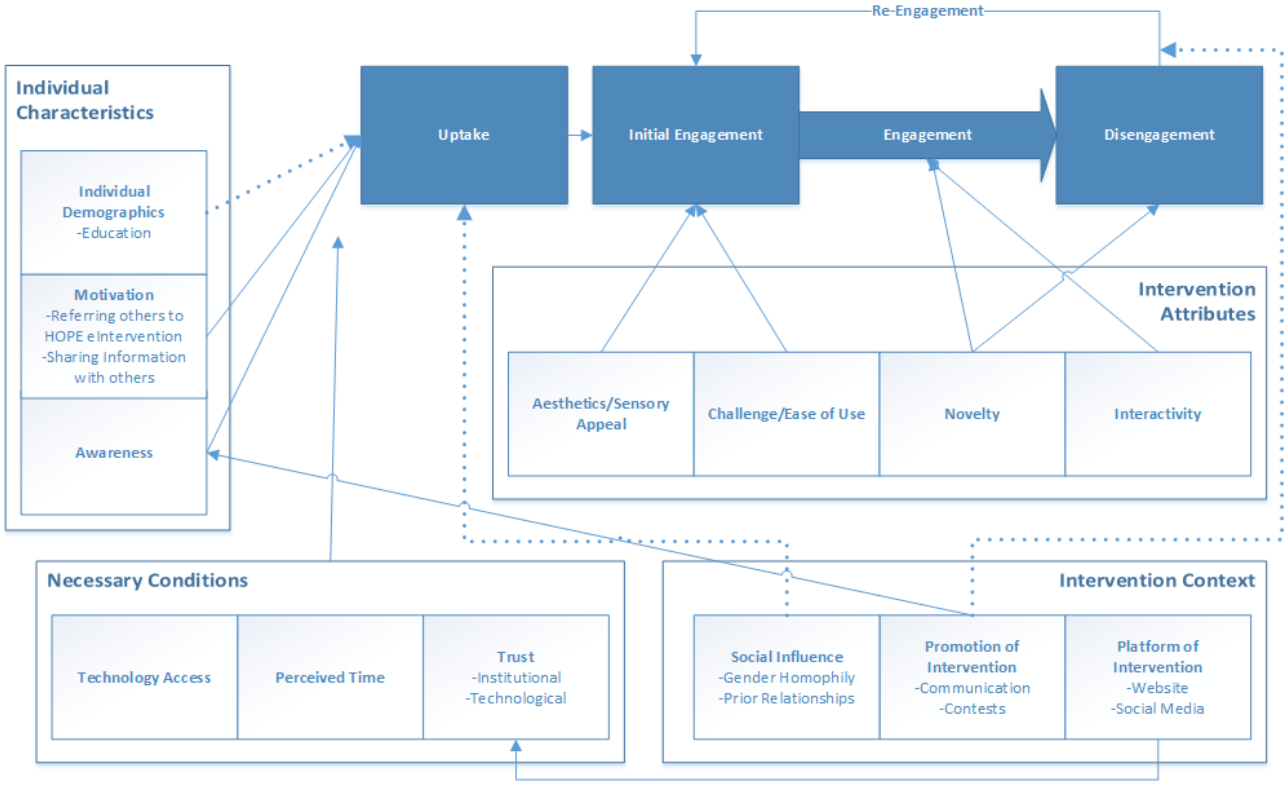
This study identified factors promoting uptake and engagement with a CHI intervention for HIV/STI prevention among African American young adults. Uptake and ongoing usage were low overall; website uptake increased as more participants entered the study (Table 3), but always remained low. The only marginally significant individual-level positive predictor of uptake was education. There was also a marginally significant positive relationship between party gender homophily and party-level uptake. Awareness and motivation to share information with others positively influenced uptake. Necessary conditions undermined uptake when absent; these included technology access, perceived lack of time, and technological trust, especially regarding social media–based discussions about HIV/STIs. Visual appeal of the website, information with the appropriate level of challenge, and interactivity positively affected ongoing usage, although the website was not interactive enough for some participants. A social media contest also increased re-engagement (a component of ongoing usage), with limited reach. Lack of novelty was linked to disengagement (another component of ongoing usage).

Comparison With Prior Work

Based on these findings, we extended O'Brien and Toms' model of engagement [31] in a "Model of CHI Intervention Uptake and Engagement for African American Young Adults" (Figure 3). Our model reinforces O'Brien and Toms' inclusion of motivation, challenge/ease of use, esthetics/sensory appeal, and novelty as factors influencing engagement; we found this in our interviews. To make the model more appropriate for this group, we added uptake. Factors driving uptake were identified through interviews (awareness and motivation) and survey data (education level and party network homophily). The model also newly incorporated necessary conditions and intervention context, including promotion efforts and the intervention's platform on social media, which was linked to technological trust. Multimedia Appendix 7 presents which details support each element of the new model in Figure 3. With these additions, we offer interventionists a framework to identify culturally relevant factors for CHI design and implementation among African American young adults.

XSL•FO
RenderX

Figure 3. Uptake and engagement model for African American young adults. The dotted lines denote marginal significance. HOPE: HIV Outreach, Prevention, and Education.



Of the three social influence variables studied, gender homophily weakly predicted uptake, aligning with the DOI theory. This comports with studies that found that network homophily influenced technology use [69,70] and increased adoption of new health behaviors in an online intervention [46]. Strong tie proportion and network density did not influence uptake, contrary to previous findings [47,48]. In prior work, strong ties and network density often resulted in the formation of subjective norms around technology use [48]. With the HOPE eIntervention, it is possible that no strong subjective norms existed. This could explain the insignificant results of the other social influence measures [71]. Future CHI intervention studies should promote normative behaviors regarding use to drive uptake.

Sharing HIV/STI information within social networks motivated uptake and engagement. This is consistent with the design of the parties as educational events within pre-existing social networks and our use of social media contests, as well as with prior work showing motivation to share health information within African American populations [72,73]. African American young adults often have limited access to HIV/STI information [45,74]. Sharing within social networks may help fill this gap, and this could be leveraged more effectively in future work. These results suggest that CHI interventions that target prevention may need to provide easily sharable information [75] applicable to the lives of the audience.

Trust in information sources is fundamental for African American technology users [12,76,77]. HOPE users' trust in YOUR Center had an effect on uptake of the HOPE parties, which in turn affected eIntervention uptake. The website incorporated design elements aligned with recommendations posited by a meta-analysis [78] to influence trust, such as lack of login, in order to protect privacy [74]. CHI interventions for this population must incorporate trust-related considerations, including privacy and security [79].

While the website was designed to protect privacy [12], uptake may have been affected by participants' prior experiences with the discussion of HIV/STI information online, and other trust-related concerns observed in prior work. Observing negative interactions shapes how individuals engage with online interventions [80] and manage online identities [81]. Therefore, these experiences might have particularly deterred use of interactive intervention elements, such as commenting on the blog. Furthermore, interactions, such as *liking* a Facebook page or *following* a Twitter account, may have been identifying [82] enough to discourage doing so. While African American young adults may believe that social media can helpfully disseminate health information [83], HIV/STI information may be too sensitive to discuss publicly.

Participants expressed interest in increased interactivity and similarity to sites such as Facebook. Modeling interventions after social media sites may promote engagement. Young adults have the highest social media usage among age groups [84] and may expect interventions to mirror their interactivity. Notably, although a blog provided interactivity on the website, decreasing use of blogs among youth over time [85] indicates that blogs may not have had the requisite popularity with the audience to provide them with desired interactivity. Additionally, given the low uptake of social media–based HOPE accounts, websites targeting HIV prevention or other health issues for youth could implement other components, such as newsfeeds, thus allowing for passive engagement [86,87].

XSL•FO
RenderX

The website fostered initial engagement with esthetically pleasing visuals and audience-appropriate information. Webpage attractiveness may be as important as content for engaging African American young adults [17], who prefer attention-getting colors [88]. Perceived information quality and ease of use also drove engagement. CHI interventions targeting specific populations should present content that is culturally relevant, accessible, and perceived as trustworthy, such as through links to social networks and trusted institutions [75].

Others have shown that lack of time and technology access may be uptake barriers [29,37]. Many survey respondents said that they lacked time to use the eIntervention. More advantaged participants are more likely to have time to engage with health promotion efforts [89], so young African Americans may particularly need CHI designs with low time burden. Additionally, a significant minority of party participants said they did not use the eIntervention because they lacked technology access. African Americans are less likely than white individuals to own computers or have broadband and are more likely to be "smartphone only" internet users [90]. Interventions targeting African American young adults or other groups that may not have access to consistent home broadband should be optimized for cellphones.

Despite attempts to increase uptake and re-engagement through face-to-face introductions to the eIntervention at HOPE parties and social media contests with financial incentives, low uptake may have resulted from insufficient promotion or limited usefulness [91]. Research on African American health intervention use found that face-to-face interactions have a greater impact on health outcomes [92]. Including more in-person promotional campaigns before implementing an intervention may be helpful.

Limitations

This study has some limitations. African American young adults were sampled from one county in one state. While we cannot generalize the results to broader populations, they could be extended to groups at a high risk for acquiring HIV/STIs in other urban settings in the United States. The survey and interview relied on participant self-report, which may have desirability effects; however, this mixed-methods approach allowed for diverse data to inform our analysis. Large amounts of data were missing in the self-report questionnaires, largely due to difficulties with follow-up despite attempts using listed phone numbers and mailing addresses, which is a common obstacle in community-engaged research. Additionally, while Google Analytics, Facebook, and Twitter data provided descriptive information about website use, we could not ensure that individual visitors were party participants because we did not implement user accounts due to privacy concerns and the website was accessible via public search [12]. Additionally, many of the respondents reported being unemployed at the point of the baseline survey, but they may have begun employment or schooling during the follow-up period.

Conclusion

Our study identified factors driving uptake and engagement within an HIV-prevention CHI intervention with African American young adults. We affirm and extend O'Brien and Toms' model to include uptake, individual factors, necessary conditions, and context as displayed in Figure 3. Findings revealed that CHI interventions for prevention among African American young adults should facilitate peer information sharing, be implemented through trusted organizations, be interactive, and offer attention-grabbing esthetic designs, accessible information, and regular updates. Intervention design and implementation must also foster trust and address barriers such as lack of time.

Acknowledgments

This research was supported by the Prevention Research Center of Michigan cooperative agreement (number 1-U48-DP-001901) from the Centers for Disease Control and Prevention. Additional funding was provided by the University of Michigan's Rackham Graduate School Spring/Summer Research Grants Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
HIV Outreach, Prevention, and Education (HOPE) eIntervention components: website.
[PNG File , 392 KB - jmir_v23i7e22203_app1.png]

Multimedia Appendix 2
HIV Outreach, Prevention, and Education (HOPE) eIntervention components: Twitter.
[PNG File , 46 KB - jmir_v23i7e22203_app2.png]

Multimedia Appendix 3
HIV Outreach, Prevention, and Education (HOPE) eIntervention components: Facebook post.
[PNG File , 537 KB - jmir_v23i7e22203_app3.png]

XSL•FO
RenderX

Multimedia Appendix 4
Survey instrument.
[DOC File , 105 KB - jmir_v23i7e22203_app4.doc]

Multimedia Appendix 5
Timeline of community partnership, intervention development, HOPE parties, and data collection. HOPE: HIV Outreach, Prevention, and Education.
[PNG File , 48 KB - jmir_v23i7e22203_app5.png]

Multimedia Appendix 6
Extended model of user engagement based on qualitative interview data.
[DOCX File , 16 KB - jmir_v23i7e22203_app6.docx]

Multimedia Appendix 7
Support for each element of the proposed model.
[DOCX File , 15 KB - jmir_v23i7e22203_app7.docx]

References

1. HIV and African American People. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/group/racialethnic/africanamericans/index.html [accessed 2021-06-29]
2. HIV and Youth. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hiv/pdf/group/age/youth/cdc-hiv-youth.pdf [accessed 2021-06-29]
3. Sexually Transmitted Disease Surveillance 2017. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/std/stats17/2017-STD-Surveillance-Report_CDC-clearance-9.10.18.pdf [accessed 2021-06-29]
4. LeGrand S, Muessig KE, Horvath KJ, Rosengren AL, Hightow-Weidman LB. Using technology to support HIV self-testing among MSM. Curr Opin HIV AIDS 2017 Sep;12(5):425-431 [FREE Full text] [doi: 10.1097/COH.0000000000000400] [Medline: 28617712]
5. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
6. Lyles C, Harris L, Jordan L, Grothaus L, Wehnes L, Reid R, et al. Patient race/ethnicity and shared medical record use among diabetes patients. Med Care 2012 May;50(5):434-440. [doi: 10.1097/MLR.0b013e318249d81b] [Medline: 22354209]
7. Ragsdale A, Rotheram-Borus M. Re-shaping HIV Interventions with Technology. AIDS Behav 2015 Jun;19 Suppl 2:77-80 [FREE Full text] [doi: 10.1007/s10461-015-1086-y] [Medline: 26088905]
8. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
9. Danielson C, McCauley J, Gros K, Jones A, Barr S, Borkman A, et al. SiHLEWeb.com: Development and usability testing of an evidence-based HIV prevention website for female African-American adolescents. Health Informatics J 2016 Jun;22(2):194-208 [FREE Full text] [doi: 10.1177/1460458214544048] [Medline: 25167865]
10. Holloway IW, Winder TJ, Lea CH, Tan D, Boyd D, Novak D. Technology Use and Preferences for Mobile Phone-Based HIV Prevention and Treatment Among Black Young Men Who Have Sex With Men: Exploratory Research. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e46 [FREE Full text] [doi: 10.2196/mhealth.6436] [Medline: 28408360]
11. Patel VV, Masyukova M, Sutton D, Horvath KJ. Social Media Use and HIV-Related Risk Behaviors in Young Black and Latino Gay and Bi Men and Transgender Individuals in New York City: Implications for Online Interventions. J Urban Health 2016 Apr;93(2):388-399 [FREE Full text] [doi: 10.1007/s11524-016-0025-1] [Medline: 26936854]
12. Veinot TC, Campbell TR, Kruger DJ, Grodzinski A. A question of trust: user-centered design requirements for an informatics intervention to promote the sexual health of African-American youth. J Am Med Inform Assoc 2013;20(4):758-765 [FREE Full text] [doi: 10.1136/amiajnl-2012-001361] [Medline: 23512830]
13. Voisin DR, Tan K, DiClemente RJ. A longitudinal examination of sexually transmitted infection/HIV prevention knowledge and sexually transmitted infections among African-American adolescent females. J Health Psychol 2013 Dec;18(12):1582-1587. [doi: 10.1177/1359105312465916] [Medline: 23221618]
14. Young LE, Schumm P, Alon L, Bouris A, Ferreira M, Hill B, et al. PrEP Chicago: A randomized controlled peer change agent intervention to promote the adoption of pre-exposure prophylaxis for HIV prevention among young Black men who have sex with men. Clin Trials 2018 Feb;15(1):44-52 [FREE Full text] [doi: 10.1177/1740774517730012] [Medline: 28862483]
15. Voisin DR, Tan K, Salazar LF, Crosby R, DiClemente RJ. Correlates of sexually transmitted infection prevention knowledge among African American girls. J Adolesc Health 2012 Aug;51(2):197-199 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.12.022] [Medline: 22824452]

XSL•FO
RenderX

16. DiClemente R, Davis T, Swartzendruber A, Fasula A, Boyce L, Gelaude D, et al. Efficacy of an HIV/STI sexual risk-reduction intervention for African American adolescent girls in juvenile detention centers: a randomized controlled trial. Women Health 2014;54(8):726-749 [FREE Full text] [doi: 10.1080/03630242.2014.932893] [Medline: 25190056]
17. Downs J, Murray P, Bruine de Bruin W, Penrose J, Palmgren C, Fischhoff B. Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized controlled trial. Soc Sci Med 2004 Oct;59(8):1561-1572. [doi: 10.1016/j.socscimed.2004.01.032] [Medline: 15279915]
18. Bond K, Ramos S. Utilization of an Animated Electronic Health Video to Increase Knowledge of Post- and Pre-Exposure Prophylaxis for HIV Among African American Women: Nationwide Cross-Sectional Survey. JMIR Form Res 2019 May 29;3(2):e9995 [FREE Full text] [doi: 10.2196/formative.9995] [Medline: 31144667]
19. Wingood GM, Card JJ, Er D, Solomon J, Braxton N, Lang D, et al. Preliminary efficacy of a computer-based HIV intervention for African-American women. Psychol Health 2011 Feb;26(2):223-234. [doi: 10.1080/08870446.2011.531576] [Medline: 21318931]
20. Kalichman SC, Cherry C, Cain D, Pope H, Kalichman M, Eaton L, et al. Internet-based health information consumer skills intervention for people living with HIV/AIDS. J Consult Clin Psychol 2006 Jun;74(3):545-554. [doi: 10.1037/0022-006X.74.3.545] [Medline: 16822111]
21. Teadt S, Burns J, Montgomery T, Darbes L. African American Adolescents and Young Adults, New Media, and Sexual Health: Scoping Review. JMIR Mhealth Uhealth 2020 Oct 05;8(10):e19459 [FREE Full text] [doi: 10.2196/19459] [Medline: 33016890]
22. Skovdal M. Facilitating engagement with PrEP and other HIV prevention technologies through practice-based combination prevention. J Int AIDS Soc 2019 Jul;22 Suppl 4:e25294 [FREE Full text] [doi: 10.1002/jia2.25294] [Medline: 31328412]
23. St Lawrence J, Eldridge G, Reitman D, Little C, Shelby M, Brasfield T. Factors influencing condom use among African American women: implications for risk reduction interventions. Am J Community Psychol 1998 Feb;26(1):7-28. [doi: 10.1023/a:1021877906707] [Medline: 9574496]
24. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/AJPH.2014.301992] [Medline: 25033137]
25. St Lawrence J, Brasfield T, Jefferson K, Alleyne E, O'Bannon R, Shirley A. Cognitive-behavioral intervention to reduce African American adolescents' risk for HIV infection. J Consult Clin Psychol 1995 Apr;63(2):221-237. [doi: 10.1037//0022-006x.63.2.221] [Medline: 7751483]
26. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019 Apr;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
27. Emery J, Coleman T, Sutton S, Cooper S, Leonardi-Bee J, Jones M, et al. Uptake of Tailored Text Message Smoking Cessation Support in Pregnancy When Advertised on the Internet (MiQuit): Observational Study. J Med Internet Res 2018 Apr 19;20(4):e146 [FREE Full text] [doi: 10.2196/jmir.8525] [Medline: 29674308]
28. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
29. Price M, Gros DF, McCauley JL, Gros KS, Ruggiero KJ. Nonuse and dropout attrition for a web-based mental health intervention delivered in a post-disaster context. Psychiatry 2012;75(3):267-284 [FREE Full text] [doi: 10.1521/psyc.2012.75.3.267] [Medline: 22913502]
30. Alkhaldi G, Hamilton F, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
31. O'Brien H, Toms E. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
32. Christopher Gibbons M. Use of health information technology among racial and ethnic underserved communities. Perspect Health Inf Manag 2011 Jan 01;8:1f [FREE Full text] [Medline: 21307989]
33. Dancy BL, Wilbur J, Talashek M, Bonner G, Barnes-Boyd C. Community-based research: barriers to recruitment of African Americans. Nurs Outlook 2004;52(5):234-240. [doi: 10.1016/j.outlook.2004.04.012] [Medline: 15499312]
34. Oh A, Chou W, Jackson D, Cykert S, Jones N, Schaal J. Chapter 2 - Reducing Cancer Disparities Through Community Engagement: The Promise of Informatics. In: Hesse BW, Ahern DK, Beckjord E, editors. Oncology Informatics: Using Health Information Technology to Improve Processes and Outcomes in Cancer. Cambridge, MA: Academic Press; 2016:23-39.
35. Broadhead RS, Heckathorn DD, Weakliem DL, Anthony DL, Madray H, Mills RJ, et al. Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. Public Health Rep 1998 Jun;113 Suppl 1:42-57 [FREE Full text] [Medline: 9722809]
36. Latkin CA, Sherman S, Knowlton A. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. Health Psychol 2003 Jul;22(4):332-339. [doi: 10.1037/0278-6133.22.4.332] [Medline: 12940388]

XSL•FO
RenderX

37. Danielson CK, McCauley JL, Jones AM, Borkman AL, Miller S, Ruggiero KJ. Feasibility of delivering evidence-based HIV/STI prevention programming to a community sample of African American teen girls via the internet. AIDS Educ Prev 2013 Oct;25(5):394-404 [FREE Full text] [doi: 10.1521/aeap.2013.25.5.394] [Medline: 24059877]
38. Kelders SM. Understanding adherence to web-based interventions. University of Twente. Enschede: University of Twente; 2012. URL: https://research.utwente.nl/en/publications/understanding-adherence-to-web-based-interventions [accessed 2021-07-07]
39. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
40. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920. [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
41. Marsch LA, Grabinski MJ, Bickel WK, Desrosiers A, Guarino H, Muehlbach B, et al. Computer-assisted HIV prevention for youth with substance use disorders. Subst Use Misuse 2011;46(1):46-56 [FREE Full text] [doi: 10.3109/10826084.2011.521088] [Medline: 21190405]
42. Rogers EM. Diffusion of Innovations. Fifth Edition. New York, NY: Free Press; 2003.
43. Singhal A, Rogers E. Combating AIDS: Communication strategies in action. Thousand Oaks, CA: Sage; 2003.
44. Unertl K, Schaefbauer C, Campbell T, Senteio C, Siek K, Bakken S, et al. Integrating community-based participatory research and informatics approaches to improve the engagement and health of underserved populations. J Am Med Inform Assoc 2016 Jan;23(1):60-73 [FREE Full text] [doi: 10.1093/jamia/ocv094] [Medline: 26228766]
45. Kimmel A, Williams T, Veinot T, Campbell B, Campbell T, Valacak M, et al. 'I make sure I am safe and I make sure I have myself in every way possible': African-American youth perspectives on sexuality education. Sex Educ 2013 Mar 01;13(2):172-185 [FREE Full text] [doi: 10.1080/14681811.2012.709840] [Medline: 23585729]
46. Centola D. An experimental study of homophily in the adoption of health behavior. Science 2011 Dec 02;334(6060):1269-1272 [FREE Full text] [doi: 10.1126/science.1207055] [Medline: 22144624]
47. Haythornthwaite C. Online Personal Networks. New Media & Society 2016 Jun 30;2(2):195-226. [doi: 10.1177/14614440022225779]
48. ten Kate S, Haverkamp S, Mahmood F, Feldberg F. Social network influences on technology acceptance: a matter of tie strength, centrality and density. In: BLED 2010 Proceeding. 2010 Presented at: Bled eConference; June 20-23, 2010; Bled, Slovenia p. 18-32 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1001&context=bled2010
49. Gubrium J, Holstein J, Marvasti A, McKinney K. The SAGE Handbook of Interview Research: The Complexity of the Craft. Thousand Oaks, CA: Sage; 2012.
50. Glaser B, Strauss A. Theoretical Sampling. In: Denzin NK, editor. Sociological Methods. New York, NY: Routledge; 2017:105-114.
51. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
52. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
53. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: A meta-analysis. Computers & Education 2017 Feb;105:1-13. [doi: 10.1016/j.compedu.2016.11.003]
54. James DCS, Harville C. Barriers and Motivators to Participating in mHealth Research Among African American Men. Am J Mens Health 2017 Nov;11(6):1605-1613 [FREE Full text] [doi: 10.1177/1557988315620276] [Medline: 26634861]
55. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
56. Wimble M. Understanding Health and Health-Related Behavior of Users of Internet Health Information. Telemed J E Health 2016 Oct;22(10):809-815. [doi: 10.1089/tmj.2015.0267] [Medline: 27045569]
57. Veinot TC, Caldwell E, Loveluck J, Arnold MP, Bauermeister J. HIV Testing Behavior and Social Network Characteristics and Functions Among Young Men Who have Sex with Men (YMSM) in Metropolitan Detroit. AIDS Behav 2016 Nov 2;20(11):2739-2761 [FREE Full text] [doi: 10.1007/s10461-016-1296-y] [Medline: 26837634]
58. Krackhardt D, Stern RN. Informal Networks and Organizational Crises: An Experimental Simulation. Social Psychology Quarterly 1988 Jun;51(2):123-140. [doi: 10.2307/2786835]
59. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press; 1994.
60. Cook SH, Bauermeister JA, Gordon-Messer D, Zimmerman MA. Online network influences on emerging adults' alcohol and drug use. J Youth Adolesc 2013 Nov 2;42(11):1674-1686 [FREE Full text] [doi: 10.1007/s10964-012-9869-1] [Medline: 23212348]
61. Haynie DL. Delinquent Peers Revisited: Does Network Structure Matter? American Journal of Sociology 2001 Jan;106(4):1013-1057. [doi: 10.1086/320298]

XSL•FO
RenderX

62. Smith AMA, Grierson J, Wain D, Pitts M, Pattison P. Associations between the sexual behaviour of men who have sex with men and the structure and composition of their social networks. Sex Transm Infect 2004 Dec;80(6):455-458 [FREE Full text] [doi: 10.1136/sti.2004.010355] [Medline: 15572613]
63. Boase J, Horrigan JB, Wellman B, Rainie L. The strength of internet ties. Washington, DC: Pew Internet & American Life Project; 2006. URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=9C803A50126A5C2DC78608CD8FBCA4B4?doi=10.1.1.212.9474&rep=rep1&type=pdf [accessed 2021-06-29]
64. Billy JO, Udry JR. The influence of male and female best friends on adolescent sexual behavior. Adolescence 1985;20(77):21-32. [Medline: 3984808]
65. Choi K, Ayala G, Paul J, Boylan R, Gregorich S. Social network characteristics and HIV risk among African American, Asian/Pacific Islander, and Latino men who have sex with men. J Acquir Immune Defic Syndr 2013 Dec 15;64(5):496-501 [FREE Full text] [doi: 10.1097/QAI.0b013e3182a7ee52] [Medline: 23933767]
66. Walter HJ, Vaughan RD, Gladis MM, Ragin DF, Kasen S, Cohall AT. Factors associated with AIDS risk behaviors among high school students in an AIDS epicenter. Am J Public Health 1992 Apr;82(4):528-532. [doi: 10.2105/ajph.82.4.528] [Medline: 1546769]
67. McCulloch C, Neuhaus J. Generalized Linear Mixed Models. In: Encyclopedia of Biostatistics. Hoboken, NJ: Wiley; 2005:1-5.
68. Saldaña J. The Coding Manual for Qualitative Researchers. Thousand Oaks, CA: Sage; 2015.
69. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci U S A 2009 Dec 22;106(51):21544-21549 [FREE Full text] [doi: 10.1073/pnas.0908800106] [Medline: 20007780]
70. Park SY. An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. Educational Technology & Society 2009;12(3):150-162 [FREE Full text]
71. Schepers J, Wetzels M. A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. Information & Management 2007 Jan;44(1):90-103. [doi: 10.1016/j.im.2006.10.007]
72. Jones LM, Wright KD, Wallace MK, Veinot TC. "Take an opportunity whenever you get it": Information Sharing among African-American Women with Hypertension. J Assoc Inf Sci Technol 2018 Jan;69(1):168-171 [FREE Full text] [doi: 10.1002/asi.23923] [Medline: 29546230]
73. Veinot TC. Interactive acquisition and sharing: Understanding the dynamics of HIV/AIDS information networks. J. Am. Soc. Inf. Sci 2009 Nov;60(11):2313-2332. [doi: 10.1002/asi.21151]
74. Guilamo-Ramos V, Lee JJ, Kantor LM, Levine DS, Baum S, Johnsen J. Potential for using online and mobile education with parents and adolescents to impact sexual and reproductive health. Prev Sci 2015 Jan;16(1):53-60. [doi: 10.1007/s11121-014-0469-z] [Medline: 24522898]
75. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
76. Paige SR, Krieger JL, Stellefson ML. The Influence of eHealth Literacy on Perceived Trust in Online Health Communication Channels and Sources. J Health Commun 2017 Jan;22(1):53-65 [FREE Full text] [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
77. Richardson A, Allen JA, Xiao H, Vallone D. Effects of race/ethnicity and socioeconomic status on health information-seeking, confidence, and trust. J Health Care Poor Underserved 2012 Nov;23(4):1477-1493. [doi: 10.1353/hpu.2012.0181] [Medline: 23698662]
78. Vega LC, Montague E, Dehart T. Trust between patients and health websites: a review of the literature and derived outcomes from empirical studies. Health Technol (Berl) 2011 Nov 18;1(2-4):71-80 [FREE Full text] [doi: 10.1007/s12553-011-0010-3] [Medline: 22288026]
79. Gibbons MC, Fleisher L, Slamon RE, Bass S, Kandadai V, Beck JR. Exploring the potential of Web 2.0 to address health disparities. J Health Commun 2011;16 Suppl 1:77-89. [doi: 10.1080/10810730.2011.596916] [Medline: 21843097]
80. Veinot TC, Campbell T, Kruger D, Grodzinski A, Franzen S. Drama and danger: the opportunities and challenges of promoting youth sexual health through online social networks. AMIA Annu Symp Proc 2011;2011:1436-1445 [FREE Full text] [Medline: 22195207]
81. Evers C, Albury K, Byron P, Crawford K. Young people, social media, social network sites and sexual health communication in Australia: "this is funny, you should watch it". International Journal of Communication 2013;7(1):263-280 [FREE Full text]
82. Dwyer C. Privacy in the Age of Google and Facebook. IEEE Technol. Soc. Mag 2011;30(3):58-63. [doi: 10.1109/mts.2011.942309]
83. Klein H, Sterk CE, Elifson KW. Knowledge about HIV in a Community Sample of Urban African Americans in the South. J AIDS Clin Res 2016 Oct;7(10) [FREE Full text] [doi: 10.4172/2155-6113.1000622] [Medline: 27891291]
84. Social Media Fact Sheet. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/pi_14-01-30_factsheet_2-copy/ [accessed 2021-06-29]

XSL•FO
RenderX

85. Social Media and Young Adults. Pew Research Center. 2010. URL: https://www.pewresearch.org/internet/2010/02/03/social-media-and-young-adults/ [accessed 2021-06-29]
86. Bayer JB, Tri u P, Ellison NB. Social Media Elements, Ecologies, and Effects. Annu Rev Psychol 2020 Jan 04;71(1):471-497. [doi: 10.1146/annurev-psych-010419-050944] [Medline: 31518525]
87. Burke M, Kraut R, Marlow C. Social capital on Facebook: differentiating uses and users. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 571-580. [doi: 10.1145/1978942.1979023]
88. Lazard AJ, Pikowski J, Horrell L, Ross JC, Noar SM, Sutfin EL. Adolescents' and Young Adults' Aesthetics and Functionality Preferences for Online Tobacco Education. J Cancer Educ 2020 Apr;35(2):373-379 [FREE Full text] [doi: 10.1007/s13187-019-1475-4] [Medline: 30734263]
89. Phelan JC, Link BG. Is Racism a Fundamental Cause of Inequalities in Health? Annu. Rev. Sociol 2015 Aug 14;41(1):311-330. [doi: 10.1146/annurev-soc-073014-112305]
90. Smartphones help Blacks, Hispanics bridge some – but not all – digital gaps with whites. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2021-06-29]
91. Anhøj J, Jensen AH. Using the internet for life style changes in diet and physical activity: a feasibility study. J Med Internet Res 2004 Sep 08;6(3):e28 [FREE Full text] [doi: 10.2196/jmir.6.3.e28] [Medline: 15471754]
92. Montague E, Perchonok J. Health and wellness technology use by historically underserved health consumers: systematic review. J Med Internet Res 2012 May 31;14(3):e78 [FREE Full text] [doi: 10.2196/jmir.2095] [Medline: 22652979]

Abbreviations

CBPR: community-based participatory research
CHI: consumer health informatics
DOI: diffusion of innovations
GED: General Education Diploma
HOPE: HIV Outreach, Prevention, and Education
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by L Brewer, S Speedie, G Pu; comments to author 17.11.20; revised version received 16.04.21; accepted 16.05.21; published 16.07.21.

Please cite as:
Williamson A, Barbarin A, Campbell B, Campbell T, Franzen S, Reischl TM, Zimmerman M, Veinot TC
Uptake of and Engagement With an Online Sexual Health Intervention (HOPE eIntervention) Among African American Young Adults: Mixed Methods Study
J Med Internet Res 2021;23(7):e22203
URL: https://www.jmir.org/2021/7/e22203
doi:10.2196/22203
PMID:34269689

©Alicia Williamson, Andrea Barbarin, Bettina Campbell, Terrance Campbell, Susan Franzen, Thomas M Reischl, Marc Zimmerman, Tiffany Christine Veinot. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Brief Intervention to Increase Uptake and Adherence of an Internet-Based Program for Depression and Anxiety (Enhancing Engagement With Psychosocial Interventions): Randomized Controlled Trial

Philip J Batterham[1], PhD; Alison L Calear[1], PhD; Matthew Sunderland[2], PhD; Frances Kay-Lambkin[3], PhD; Louise M Farrer[1], PhD; Helen Christensen[4], PhD; Amelia Gulliver[1], PhD

[1]Centre for Mental Health Research, Research School of Population Health, The Australian National University, Canberra, Australia

[2]The Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, Sydney, Australia

[3]Priority Research Centre for Brain and Mental Health, University of Newcastle., Newcastle, Australia

[4]Black Dog Institute, University of New South Wales, Sydney, Australia

Corresponding Author:
Philip J Batterham, PhD
Centre for Mental Health Research
Research School of Population Health
The Australian National University
63 Eggleston Rd
Acton
Canberra, 2601
Australia
Phone: 61 02612 ext 51031
Email: philip.batterham@anu.edu.au

Abstract

Background: Psychosocial, self-guided, internet-based programs are effective in treating depression and anxiety. However, the community uptake of these programs is poor. Recent approaches to increasing engagement (defined as both uptake and adherence) in internet-based programs include brief engagement facilitation interventions (EFIs). However, these programs require evaluation to assess their efficacy.

Objective: The aims of this hybrid implementation effectiveness trial are to examine the effects of a brief internet-based EFI presented before an internet-based cognitive behavioral therapy self-help program (*myCompass 2*) in improving engagement (uptake and adherence) with that program (primary aim), assess the relative efficacy of the *myCompass 2* program, and determine whether greater engagement was associated with improved efficacy (greater reduction in depression or anxiety symptoms) relative to the control (secondary aim).

Methods: A 3-arm randomized controlled trial (N=849; recruited via social media) assessed the independent efficacy of the EFI and *myCompass 2*. The *myCompass 2* program was delivered with or without the EFI; both conditions were compared with an attention control condition. The EFI comprised brief (5 minutes), tailored audio-visual content on a series of click-through linear webpages.

Results: Uptake was high in all groups; 82.8% (703/849) of participants clicked through the intervention following the pretest survey. However, the difference in uptake between the EFI + *myCompass 2* condition (234/280, 83.6%) and the *myCompass 2* alone condition (222/285, 77.9%) was not significant (n=565; χ^2_1=29.2; *P*=.09). In addition, there was no significant difference in the proportion of participants who started any number of modules (1-14 modules) versus those who started none between the EFI + *myCompass 2* (214/565, 37.9%) and the *myCompass 2* alone (210/565, 37.2%) conditions (n=565; χ^2_1<0.1; *P*=.87). Finally, there was no significant difference between the EFI + *myCompass 2* and the *myCompass 2* alone conditions in the number of modules started (*U*=39366.50; *z*=−0.32; *P*=.75) or completed (*U*=39494.0; *z*=−0.29; *P*=.77). The *myCompass 2* program was not found to be efficacious over time for symptoms of depression ($F_{4,349.97}$=1.16; *P*=.33) or anxiety ($F_{4,445.99}$=0.12; *P*=.98).

XSL•FO
RenderX

However, planned contrasts suggested that *myCompass 2* may have been effective for participants with elevated generalized anxiety disorder symptoms ($F_{4,332.80}$=3.50; P=.01).

Conclusions: This brief internet-based EFI did not increase the uptake of or adherence to an existing internet-based program for depression and anxiety. Individuals' motivation to initiate and complete internet-based self-guided interventions is complex and remains a significant challenge for self-guided interventions.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618001565235; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375839

(J Med Internet Res 2021;23(7):e23029) doi:10.2196/23029

KEYWORDS

implementation; mental health; adherence; uptake; engagement-facilitation intervention; internet; randomized controlled trial

Introduction

Background

Implementation science focuses primarily on implementing novel evidence-based interventions in existing health care systems [1]. However, this approach does not involve people who do not engage in health services. The adoption of evidence-based interventions outside of health care settings is particularly pertinent to mental health. The burden of mental ill health and prevalence of mental health problems are high worldwide [2-5]. However, up to two-thirds of people with mental health conditions do not seek formal evidence-based treatment [6], with many citing stigma [7] and cost [8] as barriers to seeking face-to-face treatment. One proposed solution for addressing this unmet need is to offer evidence-based, low-intensity, self-guided e–mental health (E–MH) programs broadly to people in the community with mild-to-moderate symptoms [9], which could improve the efficiency of services and increase the availability of specialist care for those with more severe mental health problems [10]. However, limited studies have investigated strategies to reduce the barriers to the implementation of self-guided psychosocial programs within the wider community.

Considerable evidence has demonstrated that self-guided internet-based programs are effective in preventing and treating symptoms of common mental health problems such as depression [10,11]. Despite significant benefits to the individual and the ability to directly address perceived key barriers to accessing treatment [12], community uptake of these programs remains poor [12,13], which strongly suggests that there must be other barriers to using internet-based treatment. Research on a self-guided E–MH program for depression found that half of its unique visitors from the general community (N=194,840) did not register for the program [14], and only half of those who registered subsequently engaged in any of the program's modules. These findings indicate that those who commenced engaging in programs likely experienced barriers to adherence. Primary care research has reported even lower rates of uptake for E–MH programs of between 3% and 25% [15]. Successful targeting and reduction of barriers to uptake by even a small amount may substantially increase the number of people receiving evidence-based treatment.

Barriers to Engaging and Adhering to E–MH Treatment

There are many reported barriers to the uptake of E–MH programs [16]. These include a preference for face-to-face therapy over E–MH programs [12,17,18]; the common perception that internet-based therapies are not as effective as face-to-face therapy [17]; and concerns about issues such as data security, limited familiarity with E–MH programs, negative attitudes toward seeking help in general, or anxiety around using the internet [9,10,15-17,19]. Similarly, adherence to E–MH programs is also a significant challenge, particularly in the community outside of research settings [14,20]. There are also many proposed reasons for low levels of adherence. Some are positive, such as the individual receiving a sufficient dosage for symptom remission. Others are neutral, such as a lack of treatment needs (eg, healthy users). Negative reasons for low adherence include low motivation, a lack of perceived improvement, or the failure of the program to adequately engage the user or meet their needs or expectations [14,20-22].

Many of these attitudinal barriers are modifiable and might be offset by benefits such as increased privacy [23], high fidelity of delivery, and increased accessibility [24]. It is proposed that challenging some of these potentially modifiable barriers before an individual began an E–MH program is an implementation strategy that might increase subsequent engagement with the program [15]. Interventions based on this concept are called acceptance-facilitation interventions (AFIs). These interventions are supported by behavior change theories [25,26], which suggest that improving attitudes and social norms for the use of E–MH interventions will lead to greater acceptability and uptake [27]. AFIs comprise a brief package of information designed to target some of the noted potential barriers to program acceptability, with the ultimate goal of increasing both uptake and adherence (engagement) to these programs [15]. Two previous randomized controlled trials (RCTs) of AFIs showed improved acceptability attitudes of E–MH programs for people with chronic pain [28], diabetes [29], and depression [15]. However, all three trials measured attitudes immediately following the AFI. Conversely, a study that followed was not able to replicate this improved acceptability following the presentation of an E–MH program for chronic pain [30]. This study was unable to improve the uptake rate or adherence to the subsequent E–MH program. The authors concluded that perhaps the AFI did not influence intervention uptake and adherence because it only targeted acceptability [30]. There is a need to

XSL•FO
RenderX

develop and evaluate brief interventions that target other factors related to uptake and adherence, in addition to attitudes regarding acceptability.

Aims

Further research is clearly needed to evaluate AFIs, or more broadly, interventions that target engagement with internet-based psychosocial programs, which we label engagement facilitation interventions (EFIs). This study adopts a model of engagement [31,32] that includes both the initiation of the program (uptake) and its ongoing use (adherence). Thus, EFIs target both uptake and adherence to programs. Currently, no known studies have examined the utility of an EFI in increasing the uptake of and adherence to an existing, publicly available E–MH program. This study describes the results of a 3-armed hybrid implementation effectiveness RCT [33] evaluating the efficacy of a newly developed EFI on uptake and engagement with an existing E–MH program. The primary aim of the study is to test the effects of the EFI on uptake and engagement; the secondary aim is to test the relative efficacy of the internet-based cognitive behavioral therapy intervention with and without the EFI and test whether increased uptake and adherence were associated with greater efficacy; and the exploratory aim is to identify moderators of differential uptake, adherence, and efficacy, contingent on between-group differences in these outcomes. The EFI used in this study was developed through an iterative participatory design process with people who have lived experience of depression or anxiety to ensure that it met their needs, while accounting for key barriers to E–MH implementation identified in previous empirical studies. Given that the EFI content was designed to target both acceptability and adherence to the E–MH program, our primary hypothesis was that the EFI would increase participants' uptake of the E–MH program, defined as the initiation of at least one module of the E–MH program, and increase adherence, defined as a greater number of modules completed.

Methods

Trial Design

A 3-arm RCT, called the *Enhancing Engagement with Psychosocial Interventions* (EEPI) trial, assessed the independent efficacy of the EFI and *myCompass 2* by comparing the conditions of (1) *myCompass 2* + EFI, (2) *myCompass 2* (alone), and (3) an attention control condition.

Ethics Approval

The ethical aspects of this research were approved by The Australian National University Human Research Ethics Committee (protocol number 2018/257).

Multimedia Appendices 1 and 2 present the CONSORT (Consolidated Standards of Reporting Trials) checklists for reporting randomized trials [34].

Interventions

The EFI

The EFI comprised brief, tailored, written, and audio-visual content (approximately 5 minutes) presented to participants in the EFI condition (EFI + *myCompass 2*) on a series of click-through linear webpages. The participants in this condition viewed the EFI after they were randomized and before commencing the E–MH program for depression and anxiety (*myCompass 2*). The EFI was delivered on an internet-based platform that also housed the surveys and control group content. Figure 1 presents some examples of the EFI. The look and feel of the EFI was based on the design of the *myCompass 2* program to create a seamless flow of the EFI intervention into *myCompass 2*.

We used principles of participatory design to create the EFI, as these can improve the perceived relevance and uptake of interventions for end users [35]. We developed the EFI through a focus group study [36] of community members who had personal lived experience of depression or anxiety, or both (n=24, four groups; male=3, female=21; see the study by Gulliver et al [36] for further details on the EFI and its development). As noted in our study [36], very few males participated in these groups. Community members in the groups suggested that the EFI content should target barriers to using these programs through the provision of personalized symptom level feedback to demonstrate program needs, information about the program content, data security, program efficacy, and finally content challenging potential social norms around using E–MH programs (eg, the belief that others do not use E–MH programs). In particular, we used written and video content that was informed by theory that examines how social norms influence the acceptability of E–MH programs (eg, "online programs and apps are being increasingly used by people in the community to look after their mental health in their own time") [15,25,27].

The EFI comprised the following components:

1. Feedback about the participant's symptom levels (visual graph) and a written description of the benefits of participating in E–MH programs, tailored to symptom levels.
2. Written information about the efficacy of the E–MH program, its content, the time commitment involved, and its data security.
3. Two testimonials (presented in a single 1-minute video) outlining the benefits of E–MH programs to provide information and normalize participation in internet-based, self-guided therapy interventions.

XSL•FO
RenderX

Figure 1. Engagement facilitation intervention content example page.



myCompass 2

The E–MH program used was an updated version of the *myCompass* program that we named for this study as *myCompass 2*. *myCompass* has previously been demonstrated to reduce symptoms of depression and anxiety in two community samples (n=89; n=720) [37,38]. The *myCompass 2* program is a fully automated and interactive self-help program that is free to the end user and was designed for people with mild-to-moderate symptoms of stress, anxiety, and depression. It is delivered without therapist assistance and can be accessed privately, at any time, on a variety of platforms, including mobile phones, tablets, and computers.

The *myCompass 2* program is similar to its predecessor and contains 14 modules derived primarily from cognitive behavioral therapy, problem-solving therapy, interpersonal psychotherapy, and positive psychology. The program was delivered over 7 weeks to enable sufficient time to complete the 14 modules. The user is expected to complete two modules per week. However, symptom reduction may occur with the completion of fewer modules [38,39]. Each module takes 30-45 minutes to complete; thus, the expected commitment of the program is approximately 60-90 minutes per week. Half (7/14, 50%) of the modules provided core transdiagnostic cognitive behavioral therapy, whereas the other half provided content targeting

XSL•FO
RenderX

specific concerns related to mental health (eg, sleep). Additional interactive features are also included such as quizzes providing real-time self-monitoring of thoughts, feelings, and behaviors; self-monitoring reminders, feedback, facts, and mental health care tips or motivational statements provided via SMS text messaging or email. Similar to the original *myCompass* [38], *myCompass 2* was designed to be tailored to the user's needs. The program screens the user and then directs them to modules that are likely to be suitable for them. The clinical approaches of the original *myCompass* and *myCompass 2* programs were identical. However, the manner in which the content is delivered, the design of the program, and the user experience were all upgraded for *myCompass 2*, focusing on personalization and ease of use. This process was influenced by input from clinicians, consumers, and information technology specialists, resulting in the following key differences: (1) faster and easier sign up for users, (2) a more comprehensive user dashboard with more options for personalization, (3) modules that are more clearly displayed and easier to navigate, and finally (4) more comprehensive symptom trackers that are easier to personalize for the user.

Attention Control Condition (HealthWatch)

The attention control condition (*HealthWatch*) has previously been shown to have high credibility [40,41]. It was approximately matched to the *myCompass 2* program for the time taken to complete and comprised 14 brief modules of written information offered over 7 weeks. The information was taken from public domain health and lifestyle information and was deliberately unrelated to mental health. Module topics included *Keeping bones strong and healthy*, *Your microbes and you*, and *The power of your pancreas*. At the end of the 6-month trial period, after the follow-up data were collected, the control group was sent information about how to access the *myCompass 2* program if they wished.

Procedure

Recruitment

Participants were recruited from the general community via a social network, Facebook, from January to March 2019. As outlined in the study protocol [42], the recruitment target of at least 231 participants per condition (n=693) to meet the power requirements for our primary hypotheses (see *Outcomes and Data Analysis* section for further details), was met within 2 months. All follow-up data were collected in November 2019. Facebook was used to ensure that a broad cross-section of the community could be reached and to maintain ecological validity, as internet-based interventions are often marketed on the web directly to consumers. We set up a Facebook page to describe the study and used paid Facebook advertisements using nature imagery (eg, trees, waterfalls) that asked, "Want to learn more about your mental health? Complete a survey and 7-week online program now" along with the ethics approval information listed earlier. We also ran a concurrent advertisement targeting male Facebook users only and using typical masculine-targeted imagery (eg, images of road journeys). This was to increase male participants and address the commonly higher ratio of female to male participants in internet-based mental health trial research [43]. To increase the representation of males in the study, we continued this advertisement targeting male participants after meeting the original target sample of 693 participants after the first month of recruitment. This advertising strategy was slower and took another month; however, at that time, we increased the percentage of males in the study from 12% to 21.9%. Participants who then clicked *sign up* were directed to the information and consent page, where they were provided the key details of the study and asked to provide their consent to participate on the web.

Participants were invited to read the information sheet and consent to participate before completing the screening measures. After consenting, participants were screened using a two-stage screening process. First, they read a list of eligibility criteria that they had to endorse to be eligible for the trial. These were as follows: (1) had not previously used the *myCompass* web-based program, (2) were not currently receiving psychological therapy, (3) had not made a suicide plan in the past month, (4) had not been diagnosed with psychosis or bipolar disorder, (5) were aged 18 years or older, and (6) were currently living in Australia. The second part of the screening process involved completing the Generalized Anxiety Disorder 7-item (GAD-7) [44] and Patient Health Questionnaire-9 (PHQ-9) [45] instruments. Consistent with the approach that internet-based programs are highly suitable for those with mild-to-moderate symptoms in the community [10], potential participants were eligible if they reported current symptoms of depression, anxiety, or both in the mild-to-moderate range (score 5-14) using the screening instruments (PHQ-9 and GAD-7) described as follows. If they scored too low (0-4) on both instruments or too high (15+) on either instrument, they were not eligible. Those who were screened as not eligible at any point in the two-stage screening process were excluded from the study and provided with relevant mental health resources. All participants endorsing the suicide screening item of the PHQ-9 were provided with a prompt that asked them to telephone Lifeline, the Suicide Call Back Service, or 000 in the case of emergency.

Treatment Allocation

We delivered the trial using the digital infrastructure portal of the Black Dog Institute, Sydney, Australia. The portal allowed for computer-generated random allocation, automatic assessments, intervention materials, and reminders to be delivered seamlessly to the participants. The portal collected web use data automatically for all conditions, allowing us to assess participants' uptake and usage (adherence) of their assigned program. After completing the two-stage screening process, participants provided an email address and selected a password they could use to log in to access their assigned treatment. Participants were then randomized to one of the three conditions, using computer-based randomization stratified by general psychological distress symptom severity (as measured by the Distress Questionnaire-5 [DQ5] [46], at pretest score 5-13 vs 14-25), age (18-45 vs ≥46 years), and gender (female or male; permuted block randomization, block size of 6 within each stratum) to ensure balance across conditions. *Prefer not to answer* and *other* were categorized as the group expected to be smaller (ie, male gender and age ≥46 years). This was completed using a computerized randomization algorithm embedded in the trial portal. Those who were assigned to the

XSL•FO
RenderX

myCompass 2 conditions (conditions 1 and 2) were required to log in again when automatically redirected to the *myCompass 2* program.

Data Collection

The intervention period ran for 7 weeks. By using their email and password, participants were able to access their assigned programs as much or as little as they preferred. Through the automated system, weekly reminder emails were sent to participants to encourage them to engage with the *myCompass 2* (conditions 1 and 2) or the attention control websites (condition 3). Participants were also sent emails to complete the 7-week posttest and 6-month follow-up surveys. Participants were sent reminder emails to complete if they had not completed the questionnaire after 1 and 2 weeks. Individual participant access to surveys was closed 4 weeks after the initial email was sent to each participant.

Incentives

We emailed participants with small incentives in the form of e-gift cards for the completion of each internet-based assessment across all conditions (posttest incentive: Aus $15 [US $11.3]; follow-up incentive: Aus $25 [US $18.8]), regardless of their level of engagement with the intervention. These incentives were considered a token of appreciation for the participants' time and effort for each survey.

Blinding

The trial was double-blinded. The participants were blinded to whether they received active or attention control interventions. They were informed that they would be randomized to receive one of three programs: (1) strategies for challenging unhelpful thoughts and behaviors (*myCompass 2*), (2) education about internet-based interventions plus program (EFI + *myCompass 2*), or (3) general health and lifestyle information (attention control). They were not provided with information about which of these interventions was expected to be the most effective. Assessments were also blinded, as they were self-report. The statistician performing the analyses was also blinded to the conditions.

Outcomes and Data Analysis

Primary Hypotheses (Aim 1)

We hypothesized that uptake (initiation of at least one module) would be higher in the EFI + *myCompass 2* condition than in the *myCompass 2* alone condition (H1). We also expected greater adherence to be observed (ie, higher number of modules completed) in the EFI + *myCompass 2* condition relative to *the myCompass 2* alone condition (H2).

Secondary Hypotheses (Aim 2)

We expected that efficacy (reduction in symptoms of depression and anxiety) would be higher in the two active *myCompass 2* intervention conditions than the *HealthWatch* attention control condition at posttest and 6-month follow-up (H3). Finally, we also expected that efficacy would be higher in the EFI + *myCompass 2* condition than in the *myCompass 2* alone condition at posttest and follow-up, and this difference would be mediated by adherence to the program (H4).

Exploratory Hypotheses

We expected that uptake (H1), adherence (H2), and efficacy (H3) would be moderated by a range of sociodemographic and psychological characteristics, including gender, age, cultural and linguistic background, education, social support, symptoms of depression and anxiety, acceptability of psychosocial internet-based programs, attitudes toward professional psychological treatment, familiarity and use of technology, personality, stigma, and mental health literacy (H5). We also predicted that secondary indices of efficacy (reductions in suicidality, distress, and disability; increases in acceptability of internet-based psychosocial interventions and quality of life) would be greatest in the EFI + *myCompass 2*, followed by *myCompass 2* alone, which would also outperform the attention control condition (H6).

Power

To detect a significant difference in uptake (H1) from 50% to 65% (a conservative baseline based on previous research [47]), with conservative difference based on previous work by Ebert et al [15] with 90% power required a sample of 231 per condition (α=.05). To detect a difference in adherence (H2), assuming a small effect of f=0.19 (the estimated median effect from previous research [47]) between active conditions required a sample of 111 per condition. For the efficacy hypotheses (H3 and H4), a sample of 110 per condition was required to find an effect size of f=0.18 (based on Proudfoot et al [38]) between active conditions relative to control over the three assessment time points (baseline, post, and 6-month follow-up) with 90% power (α=.05; r=0.5; between repeated measures). We allowed up to 30% attrition from the posttest assessments. Thus, a minimum sample size of 158 per condition was required. We aimed to recruit a sample of n=693, which was based on the largest estimate of N required (n=231 per condition). This study achieved an adequate sample size (N=849) to meet the power requirements.

Measures

Demographic Characteristics

The following demographic characteristics were assessed: gender, age category, language spoken at home, level of education (primary school, some secondary school or year 10 equivalent, year 12, certificate level I-IV, diploma or associate degree, bachelor degree, graduate diploma or graduate certificate, master's degree, or doctoral degree), employment status (full-time, part-time or casual, unemployed, not working due to study or maternity leave, retirement, etc), and region or area of residence (metropolitan area, regional area, or rural or remote area). Other assessments including previous psychological treatment, mental health literacy, and stigma were also measured as per the protocol [42] but were not included here because we were not able to investigate H5 due to null findings on H1, H2, and H3.

Depression Symptoms

The PHQ-9 [45] was used to assess the symptoms of depression. This scale consists of nine items that assess the frequency of Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV symptoms of major depression during the past two

XSL•FO
RenderX

weeks. Participants rated items on a 4-point scale ranging from 0 (not at all) to 3 (nearly every day). Scores from each item were summed to produce an overall severity score, which ranges from 0 to 27. Higher scores indicate a higher severity of depression symptoms. The PHQ-9 has previously shown good sensitivity and specificity for detecting major depression in clinical and general population samples and has also demonstrated sensitivity to change over time [48]. Internal reliability using Cronbach α was .45 (N=849) at pretest. However, Cronbach α is not recommended when participants' scores fall within a restricted range [49], such as in this study where we recruited participants with symptom scores of 5-14 only for both PHQ-9 and GAD-7. In the posttest, when the range was broader (0-26), Cronbach α for the PHQ-9 was .85.

Anxiety Symptoms

We used the GAD-7 to assess anxiety symptoms. The GAD-7 scale comprises seven items that correspond to the DSM-IV and DSM-V criteria for generalized anxiety disorder [44]. Participants rated items on the same 4-point scale as the PHQ-9. Scores for each item were summed and ranges from 0 to 21, with higher scores indicating greater symptom severity. Previous research has demonstrated that the GAD-7 has good psychometric properties in general population samples, has similar properties in detecting changes to the PHQ-9, and provides accuracy compared with clinical diagnosis [48,50]. This study sample Cronbach α was .75 (N=849) at pretest, but as above, this may not be a reliable estimate. In the posttest, Cronbach α was .88 at posttest.

General Psychological Distress

We used the DQ5 [46] to measure general psychological distress. Given that DQ5 provides coverage of both depression and anxiety symptoms, for the purpose of this study, this measure was also used to stratify participants at randomization. We selected the case finding cut-off point of ≥14 (lower distress=5-13; higher distress=14-25) for stratification [46]. The DQ5 comprises five items that ask respondents to indicate the frequency of a range of distressing situations, thoughts, and feelings over the previous 30 days using a 5-point Likert-type scale (1=never, 2=rarely, 3=sometimes, 4=often, and 5=always). Scores are summed, and total scores range from 5 to 25, with higher scores indicating more severe levels of general psychological distress. Previous studies have demonstrated that DQ5 displays high internal consistency and external validity [46,51]. The current sample DQ5 Cronbach α was .72 at pretest in the restricted sample of individuals randomized into the study.

Acceptability of Internet-Based Psychosocial Programs

The acceptability of internet-based programs was assessed using items developed and compiled by Ebert et al [15], based on the Unified Theory of Acceptance and Use of Technology. This measure assesses acceptability, with each item rated on a 5-point scale ranging from *totally disagree* to *totally agree*. For example, one item is as follows:

> *If I was suffering from psychological strain such as enduring lowered mood, loss of interest and lowered energy, sleeping problems, rumination, loss of joy in life...I could imagine trying out an internet-based intervention for mental health problems.*

Scores are summed (range 4-20); higher scores indicate higher acceptability of internet-based programs. This scale has acceptable internal consistency [15]. Out of the eight total scales created by Ebert et al [15], we selected the following three scales for this study: *performance expectancy* (4 items from Wilson and Lankton [52] and Schomerus et al [53], example item: "Using an internet-based training would reduce my mental health problems"), *effort expectancy* (3 items, from the studies by Wilson et al [52] and Schomerus et al [53], example item: "Using an internet-based depression intervention would be an easy task for me"), and finally *concerns regarding data security* (2 items, developed by Ebert et al [15], example item: "When participating in an online-training I would trust, that all information I disclose would be treated in strict confidence"). All items were rated on the same 5-point scale as previously described.

Suicidal Ideation

We used five suicide-specific items from the Psychiatric Symptom Frequency scale [54] to measure suicidal ideation. Items measure suicidal ideation and suicidal behavior, using items that cover suicidal thoughts, plans, and attempts. Respondents chose yes or no to indicate whether any of these aspects of suicidal ideation or behavior were present in the previous six months; higher scores indicated higher severity of suicidal ideation and actions. Psychiatric Symptom Frequency scale suicide items display high internal reliability and validity [54]. These items were assessed at the pretest and 6-month follow-up only.

Disability or Days Out of Role

Disability and the days out of role were measured by two items. The first assessed the number of days out of role: "In the last 30 days, how many days were you totally unable to work, study, or manage your day-to-day activities because of emotional problems (such as feeling depressed or anxious)?" The second assessed days of disability: "Aside from those days, in the last 30 days, how many days were you able to work, study, or manage your day-to-day activities but had to cut back on what you did or did not get as much done as usual because of emotional problems?"

Quality of Life

We used the European Health Interview Survey Quality of Life 8-item index (EUROHIS-QOL) [55] to measure quality of life. This scale comprises eight items measuring the psychological, physical, social, and environmental components of quality of life. Two examples include, "How would you rate your quality of life?" and "How satisfied are you with your ability to perform your daily activities?" Respondents rated the items on a 5-point scale, with response categories ranging from very dissatisfied to very satisfied, very poor to very good, and not at all to completely. A total score ranging from 0 to 32 is produced by summing the scores from each item; higher scores indicate a higher perceived quality of life. Previous research has shown adequate internal consistency for EUROHIS-QOL in multiple samples [55,56].

XSL•FO
RenderX

Perceived Reasons for Nonadherence

We also asked participants at the end of the posttest survey to complete a measure of self-reported adherence by asking, "Did you complete all of the program?" For those that responded "No," we asked, "What were some of the reasons you didn't complete the program?"

Analyses

Primary Outcomes (Aim 1)

The primary outcomes were uptake and adherence to the *myCompass 2* program. Uptake (H1) was assessed as the number of individuals who accessed at least one therapeutic module of the program. We compared the rate of uptake in EFI + *myCompass 2* to that in the *myCompass 2* alone condition using a chi-square test. Adherence (H2) was assessed using a Mann-Whitney *U* test to compare the number of modules that started and completed *myCompass 2* during the intervention period of 7 weeks. We selected modules completed as they captured the dosage of the therapeutic content received. We also examined qualitative data on self-reported reasons for nonadherence using thematic analysis.

Secondary Outcomes (Aim 2)

Efficacy (H3 and H4) was assessed on the basis of reduced symptoms of depression (PHQ-9 [45]) and GAD-7 [44] at the posttest and 6-month follow-up. Comparisons of efficacy were made between participants in each of the active intervention conditions and those in the attention control condition. We analyzed the secondary efficacy outcomes using an intention-to-treat framework, using data collected from the three measurement occasions (pretest, posttest, and 6-month follow-up). We calculated the effects with IBM SPSS 26.0 for Windows (IBM Corporation) using mixed model repeated measures (MMRM) analyses of variance, conservatively estimated using unstructured covariance matrices. Mixed models techniques incorporate all available data, including data from participants who did not complete assessments at posttest or follow-up, under the assumption that data were missing at random [57].

Exploratory Outcomes

The MMRM processes described above were also used to examine the effects of the interventions on secondary efficacy outcomes (H6). The exploratory logistic regression and negative binomial regression models, and growth mixture model analyses (planned for H5) [42] did not proceed due to null findings on H1, H2, and H3.

Results

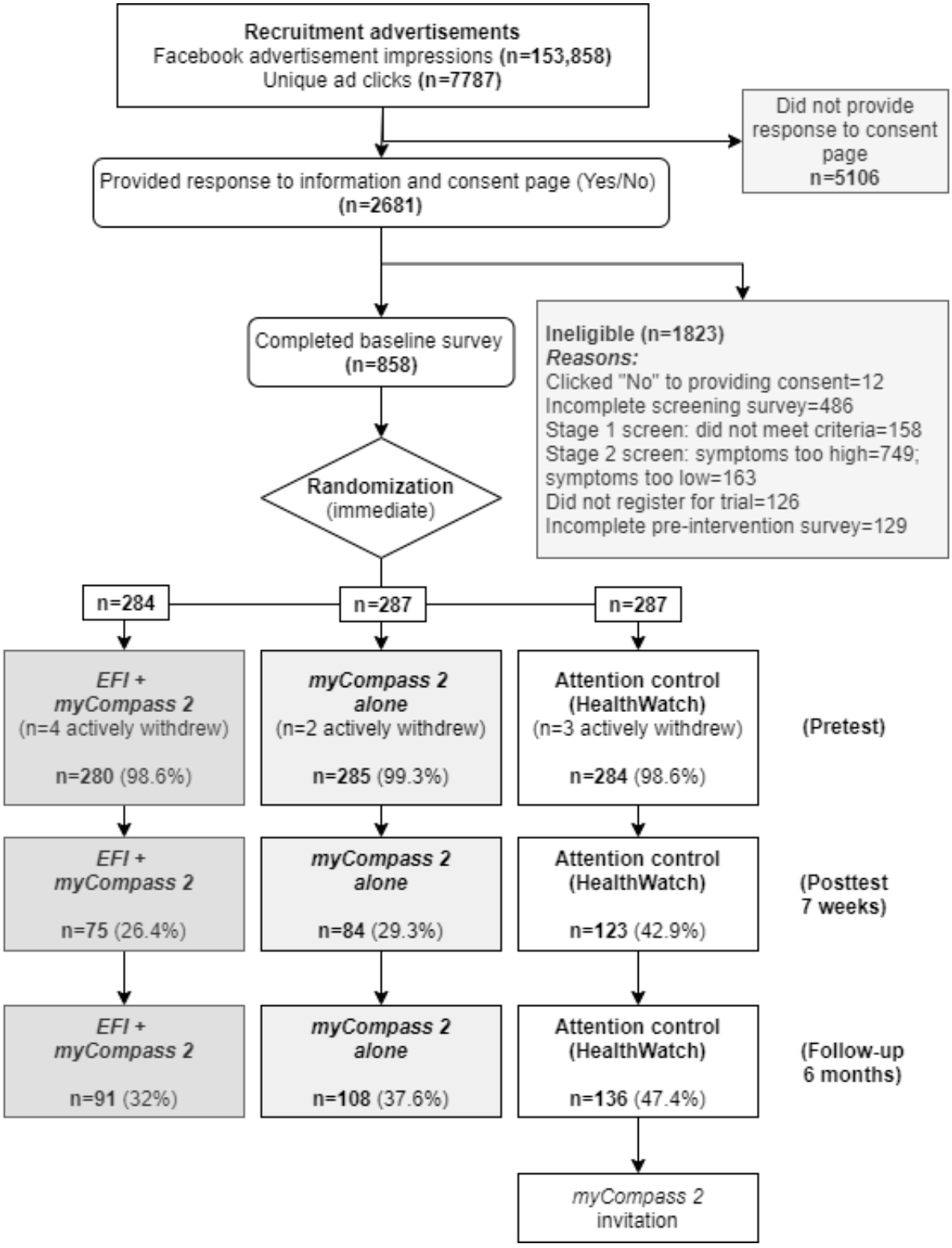
Participants

A total of 858 participants met the eligibility criteria, agreed to participate in the trial, and completed the baseline assessment. Nine participants withdrew over the course of the study, and these were evenly spread across the three conditions. Figure 2 shows a CONSORT diagram of participant flow throughout the study. Overall retention rates from randomization were 32.9% at posttest and 39% at follow-up.

Table 1 presents the participants' characteristics. There were no significant differences between the conditions at pretest for any of the characteristics.

XSL•FO
RenderX

Figure 2. Flow of participants through the trial. Percentages are from the total randomized. EFI: engagement facilitation intervention.



XSL•FO
RenderX

Table 1. Characteristics of participants included in this study (N=849).

Characteristic	Intervention arm		Attention control (n=284)	Total (N=849)
	EFI[a] + *myCompass 2* (n=280)	*myCompass 2* alone (n=285)		
Age category (years) , n (%)				
18-25	30 (10.7)	24 (8.4)	25 (8.8)	79 (9.3)
26-35	55 (19.6)	50 (17.5)	54 (19)	159 (18.7)
36-45	84 (30)	79 (27.7)	84 (29.6)	247 (29.1)
46-55	68 (24.3)	86 (30.2)	60 (21.1)	214 (25.2)
56-65	35 (12.5)	39 (13.7)	44 (15.5)	118 (13.9)
≥66	8 (2.9)	7 (2.5)	17 (6)	32 (3.8)
Gender , n (%)				
Male	57 (20.4)	61 (21.4)	68 (23.9)	186 (21.9)
Female	217 (77.5)	217 (76.1)	212 (74.6)	646 (76.1)
Other	4 (1.4)	2 (0.7)	2 (0.7)	8 (0.9)
Prefer not to answer	2 (0.7)	5 (1.8)	2 (0.7)	9 (1.1)
Highest level of education , n (%)				
High school or less	52 (18.6)	49 (17.2)	54 (19)	155 (18.3)
Certificate or diploma	108 (38.6)	93 (32.6)	103 (36.3)	304 (35.8)
Bachelor's degree	67 (23.9)	71 (24.9)	71 (25)	209 (24.6)
Postgraduate degree or diploma	51 (18.2)	72 (25.3)	55 (19.4)	178 (21)
Prefer not to answer	2 (0.7)	0 (0)	1 (0.4)	3 (0.4)
Employment , n (%)				
Full-time	96 (34.3)	96 (33.7)	100 (35.2)	292 (34.4)
Part-time or casual	99 (35.4)	83 (29.1)	91 (32)	273 (32.2)
Unemployed	35 (12.5)	44 (15.4)	29 (10.2)	108 (12.7)
Not working (eg, study or maternity leave)	46 (16.4)	56 (19.6)	59 (20.8)	161 (19)
Prefer not to answer	4 (1.4)	6 (2.1)	5 (1.8)	15 (1.8)
Language , n (%)				
English	272 (97.1)	275 (96.5)	273 (96.1)	820 (96.6)
English and other or other language only	8 (2.9)	10 (3.5)	11 (3.9)	29 (3.4)
Location , n (%)				
Metropolitan	115 (41.1)	123 (43.2)	122 (43)	360 (42.4)
Regional	132 (47.1)	119 (41.8)	121 (42.6)	372 (43.8)
Rural or remote	32 (11.4)	42 (14.7)	41 (14.4)	115 (13.5)
Prefer not to answer	1 (0.4)	1 (0.4)	0 (0)	2 (0.2)
Symptom measures, mean (SD)				
Anxiety (GAD-7[b])	7.63 (3.42)	7.28 (3.27)	7.71 (3.37)	7.54 (3.36)
Depression (PHQ-9[c])	9.82 (2.77)	9.64 (2.70)	9.49 (2.92)	9.65 (2.80)
General psychological distress (DQ5[d])	14.50 (2.96)	14.39 (3.07)	14.42 (2.87)	14.44 (2.97)
Acceptability (UTAUT[e])	14.49 (3.18)	14.42 (3.01)	14.91 (2.87)	14.61 (3.03)

XSL•FO
RenderX

Characteristic	Intervention arm		Attention control (n=284)	Total (N=849)
	EFI[a] + *myCompass 2* (n=280)	*myCompass 2* alone (n=285)		
Performance expectancy (UTAUT)	13.53 (2.67)	13.41 (2.67)	13.89 (2.70)	13.61 (2.69)
Effort expectancy (UTAUT)	10.66 (2.19)	10.67 (2.12)	10.54 (1.98)	10.62 (2.09)
Data security concerns (UTAUT)	7.48 (1.86)	7.30 (1.96)	7.37 (1.85)	7.38 (1.89)
Days out of role	3.04 (4.49)	4.14 (5.92)	3.62 (5.65)	3.61 (5.41)
Days cut down	9.01 (8.33)	9.51 (8.65)	7.60 (7.32)	8.71 (8.15)
Quality of life (EUROHIS-QOL[f])	24.19 (5.15)	23.57 (5.13)	24.46 (5.68)	24.07 (5.33)
Suicidal ideation (PSF[g])	0.96 (1.21)	0.95 (1.21)	0.84 (1.09)	0.92 (1.17)

[a]EFI: engagement facilitation intervention.

[b]GAD-7: Generalized Anxiety Disorder-7 item.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]DQ5: Distress Questionnaire-5.

[e]UTAUT: Unified Theory of Acceptance and Use of Technology.

[f]EUROHIS-QOL: European Health Interview Survey Quality of Life 8-item index.

[g]PSF: Psychiatric Symptom Frequency scale.

Missingness

A chi-square test indicated a significant difference between conditions in completion of posttest assessments (N=849; χ^2_2=19.3; *P*<.001), and 6-month follow-up assessments (N=849; χ^2_2=13.2; *P*=.001). Examination of the standardized residuals indicated that there were fewer participants missing from the control group, and a greater number of participants were missing from the EFI + *myCompass 2* condition at both the posttest and follow-up.

Uptake (H1)

Table 2 shows that overall uptake of the programs was very high (703/849, 82.8%). A chi-square analysis using Fisher's exact test demonstrated that the difference in uptake between conditions overall was significant (N=849; χ^2_2=8.3; *P*=.02). However, the difference in uptake between the EFI + *myCompass 2* (472/565, 83.6%) and the *myCompass 2* alone (440/565, 77.9%) conditions was not significant (n=565; χ^2_1=29.2; *P*=.09). There was also no significant difference in the proportion of participants who did and did not start a module in the *myCompass 2* program between the EFI + *myCompass 2* (214/565, 37.9%) and the *myCompass 2* alone (210/565, 37.2%) conditions (n=565; χ^2_1<0.1; *P*=.87).

XSL•FO
RenderX

Table 2. Uptake and adherence data (N=849).

Outcome	Intervention arm, n (%)			Total (N=849), n (%)
	EFI[a] + *myCompass 2* (n=280)	*myCompass 2* alone (n=285)	Attention control (n=284)	
Uptake (clicked through to program following pretest survey)[b]				
Yes	234 (83.6)	222 (77.9)	247 (87)	703 (82.8)
No	46 (16.4)	63 (22.1)	37 (13)	146 (17.2)
Uptake (number of modules started)				
0	174 (62.1)	179 (62.8)	36 (12.7)	389 (45.8)
1-14	106 (37.9)	106 (37.2)	248 (87.3)	460 (54.2)
1-2	69 (24.6)	74 (26)	232 (81.7)	375 (44.2)
3-6	31 (11.1)	27 (9.5)	8 (2.8)	66 (7.8)
7-14	6 (2.1)	5 (1.8)	8 (2.8)	19 (2.2)
Adherence (number of modules completed)[c]				
0	216 (77.1)	223 (78.2)	36 (12.7)	475 (55.9)
1-14	64 (22.9)	62 (21.8)	248 (87.3)	374 (44.1)
1-2	38 (13.6)	36 (12.6)	232 (81.7)	306 (36)
3-6	20 (7.1)	22 (7.7)	8 (2.8)	50 (5.9)
7-14	6 (2.1)	4 (1.4)	8 (2.8)	18 (2.1)

[a]EFI: engagement facilitation intervention.

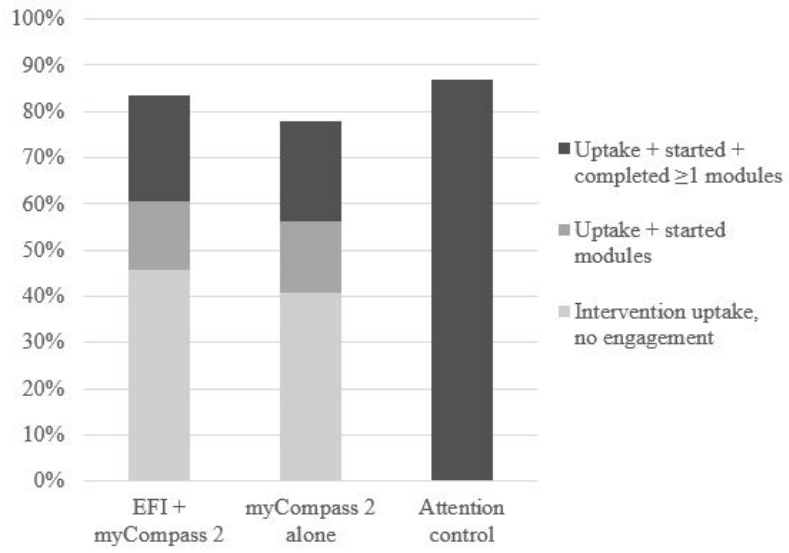
[b]One participant in the control group subsequently clicked through to the intervention 1 month after the survey was completed; thus, they are not included in the figures for uptake but are included in the modules started or completed and all their related statistical tests.

[c]Control group participants were able to access their program directly following the pretest survey without logging in, and they viewed a single page to complete a module; thus, the modules started or completed for this group are identical and not directly comparable with the intervention conditions.

Adherence (H2)

The overall adherence to the *myCompass 2* program was low. Most participants (439/565, 77.7%) across the two intervention conditions failed to complete a single module of the program. Some participants (126/565, 22.3%) completed at least part of the program, with only 1.8% (10/565) completing all 14 modules of the program. A Mann-Whitney *U* test indicated no significant difference in the number of modules started between the EFI + *myCompass 2* (mean rank=284.91) and the *myCompass 2* alone (mean rank=281.13) conditions (U=39366.50; z=−0.32; P=.75). There were no differences in the number of modules *completed* between the EFI + *myCompass 2* (mean rank=284.45) and the *myCompass 2* alone (mean rank=281.58) conditions (U=39494.0; z=−0.29; P=.77). Figure 3 presents a cumulative graph comparing the conditions of uptake and adherence to the programs.

Figure 3. Cumulative uptake and adherence to the programs by condition. EFI: engagement facilitation intervention.



XSL•FO
RenderX

Efficacy (H3 and H4)

There were no significant overall interactions between conditions and measurement occasions for depression ($F_{4,349.97}$=1.16; P=.33) or anxiety ($F_{4,445.99}$=0.12; P=.98) based on MMRM analyses. Planned contrasts demonstrated no significant interactions between time and conditions on symptoms of depression or anxiety at posttest or follow-up. As the intervention was not significantly effective overall, we restricted our planned moderation analyses of these outcomes to examine module completion, symptom levels, and basic demographics (age and gender) only. First, we examined those who completed a greater number of modules in the *myCompass 2* program. In MMRM analyses testing the effect of module completion (0 vs 1-14 modules) on mental health outcomes, the three-way interaction between module completion, time, and condition was not significant (PHQ-9: $F_{4,313.66}$=0.49; P=.74; GAD-7: $F_{4,326.46}$=0.31; P=.87), indicating no differential effects of the intervention for those who were more engaged. Table 3 provides the mean (SD) for depression and anxiety scores among the completers. We also examined the modification of intervention effects by symptom levels for both GAD-7 and PHQ-9 (none or mild vs moderate symptoms). The effects were only significant for GAD-7, suggesting that people with elevated generalized anxiety disorder symptoms at pretest may have benefitted from the intervention more than those with lower symptoms (GAD-7: $F_{4,332.80}$=3.50; P=.01). Finally, the three-way interaction between time, condition, and age (PHQ-9: $F_{4,366.09}$=1.01; P=.40; GAD-7: $F_{4,433.19}$=0.99; P=.41) and time, condition, and gender (PHQ-9: $F_{4,328.19}$=1.63; P=.17; GAD-7: $F_{4,348.97}$=1.01; P=.40) was not significant.

XSL•FO
RenderX

Table 3. Observed means and SDs for the secondary outcome measures at pre- and posttest for the trial conditions.

Measure and condition	Measurement occasion					
	Pretest		Posttest		6-month follow-up	
	Participants, n	Value, mean (SD)	Participants, n	Value, mean (SD)	Participants, n	Value, mean (SD)
Depression (PHQ-9[a])						
EFI[b] + *myCompass 2*	280	9.82 (2.77)	75	9.00 (5.18)	91	9.34 (5.47)
myCompass 2 alone	285	9.64 (2.70)	84	8.90 (5.12)	108	7.90 (4.68)
Attention control	284	9.49 (2.92)	123	8.52 (4.94)	136	8.24 (5.86)
Anxiety (GAD-7[c])						
EFI + *myCompass 2*	280	7.63 (3.42)	75	7.17 (4.49)	91	6.98 (4.56)
myCompass 2 alone	285	7.28 (3.27)	84	6.55 (4.51)	108	5.96 (4.53)
Attention control	284	7.71 (3.37)	123	6.57 (4.46)	136	6.26 (4.85)
General psychological distress (DQ5[d])						
EFI + *myCompass 2*	280	14.50 (2.96)	75	13.89 (4.15)	91	13.98 (4.16)[e]
myCompass 2 alone	285	14.39 (3.07)	81	12.96 (3.74)	105	12.50 (3.95)
Attention control	284	14.42 (2.87)	118	13.11 (3.65)	135	12.82 (4.35)
Acceptability (UTAUT[f])						
EFI + *myCompass 2*	280	14.49 (3.18)	75	14.16 (3.93)	91	12.84 (3.99)
myCompass 2 alone	285	14.42 (3.01)	81	14.42 (3.25)	105	13.51 (3.69)
Attention control	284	14.91 (2.87)	118	14.03 (3.29)	135	13.59 (3.97)
Performance expectancy (UTAUT)						
EFI + *myCompass 2*	280	13.53 (2.67)	75	13.48 (3.54)	91	12.54 (3.69)
myCompass 2 alone	285	13.41 (2.67)	81	13.86 (3.33)	105	13.14 (3.51)
Attention control	284	13.89 (2.70)	118	13.64 (3.10)	135	13.00 (3.66)
Effort expectancy (UTAUT)						
EFI + *myCompass 2*	280	10.66 (2.19)	75	10.40 (2.28)	91	10.20 (2.47)
myCompass 2 alone	285	10.67 (2.12)	81	10.30 (2.35)	105	10.12 (2.23)
Attention control	284	10.54 (1.98)	118	10.49 (2.17)	135	10.37 (2.23)
Data security concerns (UTAUT)						
EFI + *myCompass 2*	280	7.48 (1.86)	75	6.97 (1.82)	91	6.82 (2.04)
myCompass 2 alone	285	7.30 (1.96)	81	7.26 (1.99)	105	7.25 (1.93)[g]
Attention control	284	7.37 (1.85)	118	6.95 (2.07)	135	6.68 (2.16)
Days out of role						

XSL•FO
RenderX

Measure and condition	Measurement occasion					
	Pretest		Posttest		6-month follow-up	
	Participants, n	Value, mean (SD)	Participants, n	Value, mean (SD)	Participants, n	Value, mean (SD)
EFI + *myCompass 2*	280	3.04 (4.49)	75	3.81 (5.68)	91	3.52 (5.85)[h]
myCompass 2 alone	285	4.14 (5.92)	80	3.91 (5.67)	105	3.43 (6.45)
Attention control	284	3.62 (5.65)	118	3.66 (6.43)	135	2.97 (5.87)
Days cut down						
EFI + *myCompass 2*	280	9.01 (8.33)	75	9.19 (8.83)	91	9.12 (9.72)
myCompass 2 alone	285	9.51 (8.65)	80	7.13 (7.32)	105	8.55 (9.05)
Attention control	284	7.60 (7.32)	118	7.31 (8.07)	135	8.33 (9.63)
Quality of life (EUROHIS-QOL[i])						
EFI + *myCompass 2*	280	24.19 (5.15)	75	24.24 (6.63)[j]	91	24.46 (6.22)[k]
myCompass 2 alone	285	23.57 (5.13)	80	25.15 (5.81)	105	25.40 (6.06)
Attention control	284	24.46 (5.68)	118	24.86 (5.97)	135	25.75 (6.26)
Suicidal ideation (PSF[l])						
EFI + *myCompass 2*	280	0.96 (1.21)	__[m]	—	91	0.91 (1.21)
myCompass 2 alone	285	0.95 (1.21)	—	—	105	0.76 (1.13)
Attention control	284	0.84 (1.09)	—	—	135	0.82 (1.17)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]EFI: engagement facilitation intervention.

[c]GAD-7: Generalized Anxiety Disorder-7 item.

[d]DQ5: Distress Questionnaire-5.

[e]Pretest to 6-month follow-up versus *myCompass 2* alone (P=.03).

[f]UTAUT: Unified Theory of Acceptance and Use of Technology.

[g]Pretest to 6-month follow-up versus attention control (P=.01).

[h]Pretest to 6-month follow-up versus *myCompass 2* alone (P=.04) and versus attention control (P=.03).

[i]EUROHIS-QOL: European Health Interview Survey Quality of Life 8-item index.

[j]Pretest to posttest versus *myCompass 2* alone (P=.02).

[k]Pretest to 6-month follow-up versus *myCompass 2* alone (P=.02).

[l]PSF: Psychiatric Symptom Frequency scale.

[m]Data for the Psychiatric Symptom Frequency scale measures the items over the previous 6 months; Psychiatric Symptom Frequency scale data were collected at pretest and 6-month follow-up only.

Secondary Indices of Efficacy (H6)

MMRM analyses showed that there were no significant overall interactions between conditions and over time for any of the factors related to the acceptability of internet-based psychosocial interventions based on acceptance ($F_{4,316.99}$=0.39; P=.82), performance expectancy ($F_{4,343.41}$=0.55; P=.70), effort expectancy ($F_{4,357.35}$=1.31; P=.83), or concerns regarding data security ($F_{4,337.82}$=1.68; P=.16). Similarly, there were no significant effects on general psychological distress ($F_{4,382.08}$=1.31; P=.27), disability ($F_{4,361.21}$=1.70; P=.15), the days out of role ($F_{4,366.45}$=1.33; P=.26), or overall quality of life ($F_{4,357.55}$=1.99; P=.10). Suicidal ideation also did not differ between the pretest and 6-month follow-up groups ($F_{2,375.49}$=0.70; P=.50). Table 3 shows that planned contrasts demonstrated several significant interaction effects between conditions over time, although these were inconsistent across time points and none appeared to be in the expected direction.

Reasons Given for Nonadherence

A total of 47.1% (128/271) of 271 participants who responded to this question reported that they did not complete their

XSL•FO
RenderX

assigned program at posttest. This self-reported rate was similar to those who were automatically recorded to have not started (389/849, 45.8%) or completed (475/849, 55.9%) a single module of their program (Table 2). Table 4 presents the self-reported reasons for not competing with the program for the two intervention groups. Almost half (49/101, 49.5%) of participants from the *myCompass 2* alone (n=49) and EFI + *myCompass 2* (n=52) conditions reported time as an important barrier that prevented them from completing the *myCompass 2* program. Other major barriers included technical issues, simply forgetting to use the program, experiencing difficulties with concentration, or fatigue.

XSL•FO
RenderX

Table 4. Coded responses and example quotes for reasons for nonadherence given by participants in engagement facilitation intervention + *myCompass 2* and *myCompass 2* alone conditions[a].

Themes and subthemes	Example quotes (condition)	Responses coded in theme (n=144), n (%)	Respondents mentioning theme (n=101), n (%)
Structural barriers		67 (46.5)	67 (66.3)
Lack of time or competing demands	• "Just time with work and study" (EFI[b] + *myCompass 2*) • "Haven't found the time" (EFI + *myCompass 2*).	49 (34)	49 (48.5)
Technical issues	• "There were some bugs in some of the questions and answers" (*myCompass 2* alone) • "Internet at home stopped working" (*myCompass 2* alone)	18 (12.5)	18 (17.8)
Physical or mental barriers		51 (35.4)	51 (50.5)
Forgot to use program	• "Forgot. A more regular prompt would have been beneficial" (*myCompass 2* alone) • "Just forgot it was there" (EFI + *myCompass 2*)	21 (14.6)	21 (20.8)
Fatigue or concentration issues	• "I have trouble concentrating for long period of time" (EFI + *myCompass 2*) • "No mental energy" (*myCompass 2* alone)	15 (10.4)	15 (14.9)
Lack of motivation	• "It was hard to find the time and the motivation to do so" (*myCompass 2* alone) • "I didn't feel like filling it out especially if I was having a good day" (EFI + *myCompass 2*)	9 (6.3)	9 (8.9)
Too unwell	• "I was right unwell mentally and was more focus (sic) on that than internet based program." (*myCompass 2* alone)	6 (4.2)	6 (5.9)
Program barriers		16 (11.1)	16 (15.8)
Poor fit of program to needs	• "I found the information too general, not really suited to my needs and way too basic. If someone had a serious issue this would not have helped. I found I thought of it as a chore." (EFI + *myCompass 2*) • "It did not offer modules that were useful for my mental health problems" (*myCompass 2* alone)	10 (6.9)	10 (9.9)
Disliked program	• "It seemed repetitive" (EFI + *myCompass 2*). • "Bored. Activities were very samey and paint by numbers. Didn't see the connection between some activities. Needed an overview of what I should be doing and when. Couldn't see the path I was supposed to be following. Activities didn't feel specific to me." (*myCompass 2* alone)	6 (4.2)	6 (5.9)
Other		10 (6.9)	10 (9.9)
Major life events	• "Suffered an intense relationship breakdown (10 years) half way through the program" (*myCompass 2* alone)	4 (2.8)	3.9 (4)
Not accountable	• "I found it slipped to the bottom of my to-do list everyday as it did 't (sic) have a set time to do it, and no one holding me accountable" (*myCompass 2* alone)	2 (1.4)	1.9 (2)
Completed it	• "Still completing" (*myCompass 2* alone)	2 (1.4)	1.9 (2)
Miscellaneous (trust and cost)	• "Was scared of the cost" (*myCompass 2* alone)	2 (1.4)	1.9 (2)

[a]Responses were coded multiple times into themes (144 codes from 101 responses). Total data (n=101) were from participants from the *myCompass 2* alone (n=49) and EFI + *myCompass 2* (n=52) conditions only. Data from the control group were omitted (n=27).

[b]EFI: engagement facilitation intervention.

XSL•FO
RenderX

Discussion

Principal Findings

This study describes the outcomes of the EEPI trial, which involved an RCT of an EFI designed to increase uptake and adherence to a self-guided internet-based mental health program. The EFI was not found to be efficacious in improving the uptake of or adherence to E–MH intervention in this study. These findings are somewhat consistent with those of Lin et al [30], who found that despite their sample having a high acceptance of internet interventions for pain management, the uptake rate of the intervention was only moderate and adherence was very low. In contrast, uptake was high in this study, which likely reflected the minimal effort required to begin the intervention, although adherence was very low. The sample in this study was larger than that of the study by Lin et al [30], which was powered to detect more modest effects of the EFI in the context of a mental health intervention. The EFI used in this study also addressed barriers, in addition to acceptability. However, the results were similar, with no differences in uptake or adherence. The EFI in this study was also unable to significantly improve acceptability of internet interventions. The lack of difference in both uptake and acceptability may be related to ceiling effects, in that most participants were accepting of E–MH interventions and at least clicked through to the intervention. However, the lack of difference in the number of modules started suggests that EFI had minimal effect on both uptake and adherence. The findings do not preclude specific effects of the EFI; for example, some participants who received the EFI may have been motivated to engage more with the intervention, whereas others may have recognized that the intervention was not suitable for them and engaged less. However, this study shows no evidence that this implementation strategy is likely to be effective at the scale of increasing adherence.

In addition, the *myCompass 2* program was not found to be efficacious in improving depression or anxiety in this sample. In contrast, several previous RCTs have demonstrated improvements in depression and anxiety in community-based samples for the original *myCompass* program [37,38]. It is unclear whether the lack of demonstrated efficacy was related to the redesign of the *myCompass 2* program for this study, the characteristics of the sample (moderate depression or anxiety symptoms), the fully web-based nature of the trial (and consequent high attrition), or a combination of these factors. Our planned contrasts indicated that the program may be more beneficial for those with moderate anxiety symptoms. However, this result should be interpreted with caution given that the overall model for efficacy was not significant; we acknowledge that this test was post hoc and could be by chance. Further investigation is warranted to test the conditions under which this program may be effective. There is a possibility that a poor fit between the program and the needs of the participants may have had a negative impact on adherence. Additionally, the broad lack of adherence may have affected our ability to detect efficacy in this study. Nevertheless, module completion did not improve the program's efficacy, and it was not found to be effective at scale in this real-world community-based trial.

Despite the null effects of the trial, the implications of these findings remain important. There was no evidence that the implementation strategy of educating participants about their need for intervention (feedback on symptoms), benefits, perceived barriers to use, and norms of engaging with psychosocial interventions was effective in increasing engagement (uptake and adherence) with a subsequent E–MH program. Further research might evaluate whether specific components of this EFI may be able to influence the engagement of specific groups of people using factorial experiments or additional qualitative methods. It may also be the case that uptake and adherence are more challenging in the context of potentially complex mental health needs but may be more amenable to intervention in the context of other health problems [28,29].

At the outset of the study, it was clear that adherence to psychosocial interventions is a complex and dynamic behavior [14,20]. People decide not to engage with interventions for diverse reasons, many of which are entirely appropriate [14,20-22]. Targeting or tailoring both implementation strategies and interventions to the needs of an individual may be required to improve engagement. Adaptive interventions tailored to the barriers relevant to the individual, with ongoing check-ins over the duration of the intervention period may be more successful than a one-time, low-intensity strategy. This may be of benefit, as many participants simply noted *forgetting* as a barrier to using the program. Blending human support with a self-guided program may also be beneficial for increasing the uptake of internet interventions [58], although a blended approach would come at the cost of making the intervention less scalable, as human support requires additional resources. Human support for internet-based interventions may be critically important for certain individuals. An overwhelming majority of participants who noted that they did not complete the program believed that a lack of time, or competing demands for their time, was a strong barrier to completing the program. However, having accountability to another person may assist in challenging this barrier [36], as it requires commitment and time to be set aside in advance. Nevertheless, this issue remains complex, as a requirement for human contact may deter some individuals from signing up for such a program.

Importantly, the quality of the therapeutic alliance can influence treatment outcomes [59], regardless of whether it is with a computer or human. Consequently, taking into account human-computer interactions in the design of both the implementation strategy and the psychosocial intervention may also promote engagement by increasing a sense of therapeutic alliance or human connection in the internet-based setting. The co-design of interventions with end users is also imperative to ensure that interventions meet the needs and preferences of those who stand to benefit. Nevertheless, based on the current findings, our partnership with end users was not sufficient to realize the aims of the EFI. The need for psychological services will continue to increase and cannot be fully met by increasing the health professional workforce. Creative and rigorous methods to increase the use of self-guided interventions in the community, for prevention and treatment of health problems, may lead to reduction in disease burden over time.

XSL•FO
RenderX

Limitations

Although this was one of the first studies to rigorously evaluate an implementation strategy to increase engagement with a self-guided psychosocial intervention, there were some limitations of both the EFI and this study that should be noted.

Engagement Facilitation Intervention

First, as noted above, the EFI may have been too brief or insufficiently tailored to the needs of users. Although our development approach involved considerable collaboration with people who had lived experience of depression or anxiety, it is possible that the intensity of the EFI was insufficient or that it did not meet the diverse needs of users in the trial. We also did not assess participants' engagement with the EFI (eg, how much time they spent on it or if they read or watched the content, or just clicked through the pages). Attrition was the greatest in the EFI condition. It is possible that there was greater disengagement in the study in this condition, which may have occurred if the information from the EFI had a demotivating effect (ie, provided information that the psychological intervention was not of interest to the participant). When educating potential users about psychosocial interventions, there remains a risk of either overwhelming users with information or inadvertently reducing their motivation to engage.

Study Design

Overall, the attrition from the study was considerable. Although the study remained well-powered to detect hypothesized effects on uptake and adherence, it was slightly underpowered to examine efficacy outcomes, as the final samples in the active conditions were less than the targets (n=111). Attrition may have also led to biases in the analyses, although rigorous MMRM models were used to account for all available data. Incentives were used to minimize attrition but clearly provided insufficient motivation for most participants. Attrition from fully internet-based trials remains a challenge, which indicates that some form of human contact in a research trial is likely to be necessary to maintain samples over extended periods. Moreover, there were some technical challenges in the delivery of the trial. Participants could not be automatically logged into the *myCompass 2* program and were required to sign up for the intervention as a separate process, which may have led to reductions in uptake for both active conditions and raised the slight possibility of dual accounts that we may not have been able trace (ie, greater adherence than observed). The lack of efficacy of the *myCompass 2* program in this study suggests that the intervention may be better suited to participants with different symptom profiles or that further work is needed to refine the intervention based on the low adherence rates. Nevertheless, the lack of evidence for efficacy did not limit our ability to compare the levels of uptake or adherence within the RCT design of the study.

The use of Facebook or certain imagery in advertisements may have attracted a certain type of user; however, we believe this is not a significant limitation as it reflects a similar process for real-world marketing of internet-based interventions and typical users of internet-based programs, and the promotion of the trial was identical across the three arms of the trial. Participants in the trial may have been aware of their allocation based on the content, despite interventions not being explicitly labeled as active or control. Finally, the composition of the sample was biased toward female participants. Although this imbalance reflects the usage of E–MH programs in the community, it may have limited the generalizability of the results for males.

Conclusions

Although there is considerable scope for self-guided psychosocial programs to reduce health burdens, the uptake and adherence to these programs in the general population is limited. EFIs have been proposed as a specific strategy to overcome the implementation gap in psychosocial programs. However, this study indicates that the strategy was not effective in the context of an internet-based mental health program based on cognitive behavioral therapy for people with mild-to-moderate symptoms of depression or anxiety. Further research is required to identify implementation strategies that consider the dynamic and complex nature of intervention adherence and minimize the engagement barriers associated with internet-based programs.

Acknowledgments

The authors would like to thank the Black Dog Institute information technology team for their assistance in conducting the trial. This project was supported by a National Health and Medical Research Fund grant (grant 1138713). The lead author can disseminate the results of this trial without the express permission of the funder. PJB is supported by National Health and Medical Research Council (NHMRC) fellowship 1158707; ALC is supported by NHMRC fellowships 1122544 and 1173146; FKL is supported by NHMRC fellowship 1110371. LMF is supported by the ARC DECRA DE190101382. HC is supported by the NHMRC fellowship 1056964.

Conflicts of Interest

HC is the director and chief scientist at the Black Dog Institute, where the *myCompass* program was developed.

Multimedia Appendix 1
CONSORT 2010 (Consolidated Standards of Reporting Trials) checklist for reporting randomized trials.
[PDF File (Adobe PDF File), 182 KB - jmir_v23i7e23029_app1.pdf]

Multimedia Appendix 2

XSL•FO
RenderX

CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2669 KB - jmir_v23i7e23029_app2.pdf]

References

1. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
2. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013 Mar 18;34(1):119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
3. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
4. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med 2012 Jul 10;43(5):897-910. [doi: 10.1017/s003329171200147x]
5. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/s2215-0366(15)00505-2]
6. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul 1;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
7. Corrigan PW, Druss BG, Perlick DA. The impact of mental illness stigma on seeking and participating in mental health care. Psychol Sci Public Interest 2014 Oct 3;15(2):37-70. [doi: 10.1177/1529100614531398] [Medline: 26171956]
8. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999-2010. Health Aff (Millwood) 2013 Oct;32(10):1723-1730 [FREE Full text] [doi: 10.1377/hlthaff.2013.0133] [Medline: 24101061]
9. Australian Government Response to Contributing Lives, Thriving Communities: Review of Mental Health Programmes and Services. Canberra, Australia: Australian Government Department of Health; 2015.
10. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep 23;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
11. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 1;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
12. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health - attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14(1):109. [doi: 10.1186/1471-244x-14-109]
13. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
14. Batterham P, Neil A, Bennett K, Griffiths K, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adher 2008:97-105. [Medline: 19920949]
15. Ebert D, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 1;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
16. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2008 Sep 24;39(5):705-712. [doi: 10.1017/s0033291708004224]
17. Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. Ann Behav Med 2010 Aug 21;40(1):89-98 [FREE Full text] [doi: 10.1007/s12160-010-9203-7] [Medline: 20652466]
18. Casey LM, Wright M, Clough BA. Comparison of perceived barriers and treatment preferences associated with internet-based and face-to-face psychological treatment of depression. Int J Cyber Behav Psychol Learn 2014;4(4):16-22. [doi: 10.4018/ijcbpl.2014100102]
19. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun 1;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]
20. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
21. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

XSL•FO
RenderX

22. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
23. Leibert T, Archer J, Munson J, York G. An exploratory study of client perceptions of internet counseling and the therapeutic alliance. J Ment Health Counsel 2006:28. [doi: 10.17744/mehc.28.1.f0h37djrw89nv6vb]
24. Webb CA, Rosso IM, Rauch SL. Internet-based cognitive-behavioral therapy for depression: current progress and future directions. Harv Rev Psychiatry 2017;25(3):114-122. [doi: 10.1097/hrp.0000000000000139]
25. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
26. Madden TJ, Ellen PS, Ajzen I. A comparison of the theory of planned behavior and the theory of reasoned action. Pers Soc Psychol Bull 2016 Jul 2;18(1):3-9. [doi: 10.1177/0146167292181001]
27. Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
28. Baumeister H, Seifferth H, Lin J, Nowoczin L, Luking M, Ebert D. Impact of an acceptance facilitating intervention on patients' acceptance of internet-based pain interventions: a randomized controlled trial. Clin J Pain 2015;31(6):528-535. [doi: 10.1097/ajp.0000000000000118]
29. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of Internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]
30. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925]
31. Piotrowska PJ, Tully LA, Lenroot R, Kimonis E, Hawes D, Moul C, et al. Mothers, fathers, and parental systems: a conceptual model of parental engagement in programmes for child mental health-connect, attend, participate, enact (CAPE). Clin Child Fam Psychol Rev 2017 Jun 2;20(2):146-161 [FREE Full text] [doi: 10.1007/s10567-016-0219-9] [Medline: 27914017]
32. Ingoldsby EM. Review of interventions to improve family engagement and retention in parent and child mental health programs. J Child Fam Stud 2010 Oct 1;19(5):629-645 [FREE Full text] [doi: 10.1007/s10826-009-9350-2] [Medline: 20823946]
33. Curran G, Bauer M, Mittman B, Pyne J, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med care 2012;50(3):-. [doi: 10.1097/mlr.0b013e3182408812]
34. Schulz K, Altman D, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3):e1000251 [FREE Full text] [doi: 10.1371/journal.pmed.1000251] [Medline: 20352064]
35. Bovaird T. Beyond engagement and participation: user and community coproduction of public services. Public Admin Rev 2007 Sep;67(5):846-860. [doi: 10.1111/j.1540-6210.2007.00773.x]
36. Gulliver A, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Banfield M, et al. Consumer-guided development of an engagement-facilitation intervention for increasing uptake and adherence for self-guided web-based mental health programs: focus groups and online evaluation survey. JMIR Form Res 2020 Oct 29;4(10):e22528 [FREE Full text] [doi: 10.2196/22528] [Medline: 33118939]
37. Clarke J, Proudfoot J, Ma H. Mobile phone and web-based cognitive behavior therapy for depressive symptoms and mental health comorbidities in people living with diabetes: results of a feasibility study. JMIR Ment Health 2016 May 31;3(2):e23 [FREE Full text] [doi: 10.2196/mental.5131] [Medline: 27245948]
38. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13(1):312. [doi: 10.1186/1471-244x-13-312]
39. Whitton AE, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015 Mar 4;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
40. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psych 2016 Apr;3(4):333-341. [doi: 10.1016/s2215-0366(15)00536-2]
41. Christensen H, Guastella AJ, Mackinnon AJ, Griffiths KM, Eagleson C, Batterham PJ, et al. Protocol for a randomised controlled trial investigating the effectiveness of an online e-health application compared to attention placebo or sertraline in the treatment of generalised anxiety disorder. Trials 2010 Apr 30;11(1):48 [FREE Full text] [doi: 10.1186/1745-6215-11-48] [Medline: 20429947]
42. Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Gulliver A. A brief intervention to increase uptake and adherence of an online program for depression and anxiety: Protocol for the Enhancing Engagement with Psychosocial

XSL•FO
RenderX

Interventions (EEPI) Randomized Controlled Trial. Contemp Clin Trials 2019 Mar;78:107-115. [doi: 10.1016/j.cct.2019.01.015] [Medline: 30711664]
43. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
45. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. J Am Med Assoc 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
46. Batterham PJ, Sunderland M, Carragher N, Calear AL, Mackinnon AJ, Slade T. The Distress Questionnaire-5: population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016 Mar;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]
47. Kwasnicka D, Presseau J, White M, Sniehotta FF. Does planning how to cope with anticipated barriers facilitate health-related behaviour change? A systematic review. Health Psychol Rev 2013 Feb 19;7(2):129-145. [doi: 10.1080/17437199.2013.766832]
48. Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010 Jul;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
49. Fife DA, Mendoza JL, Terry R. The assessment of reliability under range restriction. Educ Psychol Measure 2012 Apr 19;72(5):862-888. [doi: 10.1177/0013164411430225]
50. Lowe B, Decker O, Muller S, Brahler E, Schellberg D, Herzog W, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/mlr.0b013e318160d093]
51. Batterham PJ, Sunderland M, Slade T, Calear AL, Carragher N. Assessing distress in the community: psychometric properties and crosswalk comparison of eight measures of psychological distress. Psychol Med 2017 Oct 2;48(8):1316-1324. [doi: 10.1017/s0033291717002835]
52. Wilson EV. Modeling patients' acceptance of provider-delivered e-health. J Am Med Info Assoc 2004 Apr 2;11(4):241-248. [doi: 10.1197/jamia.m1475]
53. Schomerus G, Matschinger H, Angermeyer MC. Attitudes that determine willingness to seek psychiatric help for depression: a representative population survey applying the Theory of Planned Behaviour. Psychol Med 2009 Apr 20;39(11):1855-1865. [doi: 10.1017/s0033291709005832]
54. Lindelow M, Hardy R, Rodgers B. Development of a scale to measure symptoms of anxiety and depression in the general UK population: the psychiatric symptom frequency scale. J Epidemiol Commun Health 1997 Oct 1;51(5):549-557 [FREE Full text] [doi: 10.1136/jech.51.5.549] [Medline: 9425466]
55. da Rocha NS, Power MJ, Bushnell DM, Fleck MP. The EUROHIS-QOL 8-item index: comparative psychometric properties to its parent WHOQOL-BREF. Value Health 2012 May;15(3):449-457 [FREE Full text] [doi: 10.1016/j.jval.2011.11.035] [Medline: 22583455]
56. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur J Public Health 2006 Aug;16(4):420-428. [doi: 10.1093/eurpub/cki155] [Medline: 16141303]
57. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. New York, USA: Springer-Verlag; 2000.
58. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
59. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 2;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]

Abbreviations

AFI: acceptance-facilitation intervention
CONSORT: Consolidated Standards of Reporting Trials
DQ5: Distress Questionnaire-5
DSM: Diagnostic and Statistical Manual of Mental Disorders
EEPI: Enhancing Engagement with Psychosocial Interventions
EFI: engagement facilitation intervention
E–MH: e–mental health
EUROHIS-QOL: EUROHIS-Quality of Life 8-item index
GAD-7: Generalized Anxiety Disorder 7-item

XSL•FO
RenderX

MMRM: mixed model repeated measures
NHMRC: National Health and Medical Research Council
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial

Edited by R Kukafka; submitted 30.07.20; peer-reviewed by J Ohan, YH Kwan, R Kok; comments to author 17.11.20; revised version received 24.11.20; accepted 04.06.21; published 27.07.21.

Please cite as:
Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Christensen H, Gulliver A
A Brief Intervention to Increase Uptake and Adherence of an Internet-Based Program for Depression and Anxiety (Enhancing Engagement With Psychosocial Interventions): Randomized Controlled Trial
J Med Internet Res 2021;23(7):e23029
URL: https://www.jmir.org/2021/7/e23029
doi:10.2196/23029
PMID:34313595

©Philip J Batterham, Alison L Calear, Matthew Sunderland, Frances Kay-Lambkin, Louise M Farrer, Helen Christensen, Amelia Gulliver. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Turn on, Tune in, and Drop out: Predictors of Attrition in a Prospective Observational Cohort Study on Psychedelic Use

Sebastian Hübner[1], BSc, MSc; Eline Haijen[1], BSc, MSc; Mendel Kaelen[1], BSc, MSc, PhD; Robin Lester Carhart-Harris[1], BSc, MA, PhD; Hannes Kettner[1], BSc, MSc

Centre for Psychedelic Research, Imperial College London, London, United Kingdom

Corresponding Author:
Hannes Kettner, BSc, MSc
Centre for Psychedelic Research
Imperial College London
Du Cane Rd
Burlington Danes Building
London, W12 0TY
United Kingdom
Phone: 44 020 7589 5111
Email: hannes.kettner17@imperial.ac.uk

Abstract

Background: The resurgence of research and public interest in the positive psychological effects of psychedelics, together with advancements in digital data collection techniques, have brought forth a new type of research design, which involves prospectively gathering large-scale naturalistic data from psychedelic users; that is, before and after the use of a psychedelic compound. A methodological limitation of such studies is their high attrition rate, particularly owing to participants who stop responding after initial study enrollment. Importantly, study dropout can introduce systematic biases that may affect the interpretability of results.

Objective: Based on a previously collected sample (baseline n=654), here we investigated potential determinants of study attrition in web-based prospective studies on psychedelic use.

Methods: Logistic regression models were used to examine demographic, psychological trait and state, and psychedelic-specific predictors of dropout. Predictors were assessed 1 week before, 1 day after, and 2 weeks after psychedelic use, with attrition being defined as noncompletion of the key endpoint 4 weeks post experience.

Results: Predictors of attrition were found among demographic variables including age (β=0.024; *P*=.007) and educational levels, as well as personality traits, specifically conscientiousness (β=–0.079; *P*=.02) and extraversion (β=0.082; *P*=.01). Contrary to prior hypotheses, neither baseline attitudes toward psychedelics nor the intensity of acute challenging experiences were predictive of dropout.

Conclusions: The baseline predictors of attrition identified here are consistent with those reported in longitudinal studies in other scientific disciplines, suggesting their transdisciplinary relevance. Moreover, the lack of an association between attrition and psychedelic advocacy or negative drug experiences in our sample contextualizes concerns about problematic biases in these and related data.

(J Med Internet Res 2021;23(7):e25973) doi:10.2196/25973

KEYWORDS

attrition; digital data; dropout; educational level; personality; psychedelics; web-based research; web-based survey

Introduction

Psychedelic substances, such as mescaline, psilocybin, or dimethyltryptamine, have likely been consumed by humans for thousands of years through different species of plant and fungi [1,2]. After a period of promising scientific exploration, especially of lysergic acid diethylamide as a psychiatric treatment aid in the 1950s and 1960s, classical psychedelics, defined here as psychoactive compounds eliciting their effects on cognition and perception through agonistic action at the serotonin 2A receptor, have been policed restrictively in most Western countries, largely prohibiting academic research [3]. The ongoing resurgence of studies into the psychological and neural effects of psychedelic substances, sometimes referred to

XSL•FO
RenderX

as a "psychedelic renaissance" [4], is now being paralleled by commercial and policy developments, reflecting a progressive social acceptance of psychedelic substances both as medicinal tools in the treatment of addictive and mood disorders [5] and for psychological benefits in healthy subjects [6].

The increasingly widespread use of psychedelics [7,8] together with advancements in digital data collection techniques, have motivated the development of a new type of ecological study focused on gathering large-scale longitudinal data sets from psychedelic users by using prospective study designs; that is, before and after the naturalistic use of a psychedelic compound. This approach has already yielded improved, ecologically valid models of the notoriously difficult-to-predict psychedelic state and its outcomes [9-11]. Efficient recruitment and data collection, low cost, and avoidance of human transcription errors represent relevant advantages of web-based survey studies such as these. However, there are also significant limitations, including the lack of experimental controls, participant accountability, and data validity, which may weaken inferences that can be drawn from the data [12,13]. Study attrition and self-selection are particular issues that could skew study samples in potentially problematic ways; for example, by promoting confirmation biases that exaggerate or downplay risks or benefits [14]. Such biases may be especially poignant when interventions that have a special value or significance for the participant are assessed [15]. Accordingly, researchers are obliged to place caveats on data collected via web-based surveys, emphasizing their preliminary, nonconfirmatory nature.

Previous studies have indicated that younger age [16-19], lower educational levels [20-26], and unemployment [24,27,28] are among the most reliable predictors of premature termination of studies or discontinuation of treatment compliance in several contexts. Psychological variables associated with poor compliance and retention include poor mental health [29-34] and low conscientiousness scores [35,36], and there is some evidence that high extraversion [37] and low agreeableness [1,38] might also be risk factors for dropout.

Although the issue of attrition in psychedelic research has not yet been addressed empirically, it has been discussed previously [9,10,39] as a significant limitation in observational psychedelic survey studies, potentially affecting the interpretability of results owing to psychedelic-specific variables associated both with psychological outcomes and the likelihood of dropout. Specifically, positive biases toward psychedelic substances have been identified in previous opportunity samples [9,40] and have in 1 case been shown to predict the increase in self-reported psychological well-being following psychedelic use [9]. A positive relationship between participant bias and the likelihood of study completion could thus represent a potential confounder leading to biased outcomes. Similarly, it is conceivable how particularly unpleasant or difficult psychedelic experiences, which are known to negatively impact long-term psychological outcomes [11], could reduce the motivation of participants to continue responding to a study, thereby creating a systematic attrition bias specific to prospective studies on psychedelics.

In this study, we used data from a published prospective assessment of the effects of psychedelic drug use on various subjective psychometric outcomes [9]. This particular study focused on identifying response predictors. The primary outcome was prospective change in psychological well-being, and consistent with a prior controlled study [41] the quality of the acute psychedelic experience was a strong predictor of longer-term outcomes, where, for instance, positive "mystical-type" experiences and high "emotional breakthrough" scores [11] were significantly predictive of improvements in psychological well-being. Problematically, however, attrition rates in this study were high, with only 29% of the total sample completing surveys up to the primary endpoint 4 weeks post psychedelic use. Therefore, this study aimed to identify variables that are most strongly associated with study noncompletion, hoping to shed light on the extent of potential attrition biases in this and similar prospective studies on psychedelics.

Statistical analyses included multiple variables and were therefore exploratory; however, we were particularly interested in the effects of positive attitudes toward psychedelic use at baseline and of difficult drug experiences on attrition, owing to their potential implications for data interpretability. Specifically, we hypothesized that study completers would have higher baseline psychedelic advocacy scores and less challenging acute drug experiences than dropouts. Based on previous studies, we also hypothesized that dropouts would be younger and have lower educational levels, employment rates, and mental health than completers.

Methods

Ethics Approval

The study was approved by Imperial College Research Ethics Committee and the Joint Research Compliance Office at Imperial College London and carried out in accordance with principles of good clinical practice. Written informed consent was obtained from all subjects. The original survey is now closed, but the revised and still active versions of related surveys are available on the PsychedelicSurvey website [42].

Design

Data were collected as part of a larger prospective study [9], approved by the Imperial College Research Ethics Committee and Joint Research Compliance Office. Only the elements of the design and data, which are relevant to this study, are presented here. Data were collected on the internet from a convenience sample of psychedelic drug users in a noncontrolled, naturalistic, and observational manner, through the website and software PsychedelicSurvey platform [42]. The open survey study was advertised on social media platforms, and informed consent was collected through a button-click feature following information on study purpose, design, and recruitment criteria. After sign-up, participants received emails that contained links to the relevant surveys at multiple timepoints, which were implemented and hosted by the web-based service system Alchemer [43]. Data were collected through a prospective cohort design: the baseline timepoint was set at 1 week before a planned psychedelic experience (timepoint 1), preacute measures were taken 1 day before the planned psychedelic experience (timepoint 2), and postacute measures were taken 1 day after the planned psychedelic experience

XSL•FO
RenderX

(timepoint 3); the first endpoint was 2 weeks after the relevant experience (timepoint 4). The subsequent key endpoint occurred 4 weeks after the planned psychedelic experience (timepoint 5).

Baseline demographic and trait variables, postacute subjective effects measures, and outcome measures at the first endpoint were used to predict attrition. The survey 4 weeks after the psychedelic experience was the key endpoint, the completion of which was used as the criterion to determine attrition vs completion of the study. Dropouts were defined as those participants who stopped responding to the study surveys and did not return to finish the key endpoint, and completers were defined as those participants who reached the key endpoint 4 weeks after the psychedelic experience, even if they missed one or more previous timepoints.

Measures

Baseline

The following measures were recorded at baseline: demographics (including age, gender, education, and employment status) and scores on the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) [44] that is used to asses psychological well-being, the 10-Item Personality Inventory [45], the Social Connectedness Scale [46], a modified version of the Tellegen Absorption Scale [47] that is used to measure trait absorption and is a reliable predictor of the intensity of psychedelic experiences [9], the short version of the Spielberger State-Trait Anxiety Inventory (STAI-SF) [48], the 16-item Quick Inventory of Depressive Symptomatology (QIDS) [49], the Suicidal Ideation Attributes Scale (SIDAS) [50], and Peters' Delusional Inventory (PDI) [51]. Additionally, 4 self-constructed items ("I am an active advocate of psychedelic drug-use," "I am an active advocate of the therapeutic use of psychedelics," "I have an advanced knowledge about psychedelics," and "I am an highly experienced psychedelic user") measured on a 5-point Likert scale were used to assess the advocacy of psychedelic drug use. A sum score was calculated on the basis of these 4 items, termed "psychedelic advocacy." Cronbach α, as a measure of internal consistency among these 4 items, indicated acceptable internal consistency at α=.78 (bootstrap 95% CI 0.75-0.81 for 1000 samples).

Postacute Timepoint

Retrospective surveys pertaining to the acute experience were answered 1 day after the relevant psychedelic experience. These included a visual effects (VE) subscale of the 11-dimension altered states of consciousness questionnaire [52]; the Mystical Experience Questionnaire [53] that assesses acute positive peak experiences; the Challenging Experience Questionnaire [54] that is a measure of unpleasant affective, cognitive, and somatic difficulties experienced during psychedelic use; the Ego Dissolution Inventory [55] that addresses reductions in self-referential processing; the Emotional Breakthrough Inventory [11]; and the Physical Symptoms Inventory [56] that helps determine the number of several physical symptoms possibly experienced postacutely.

Endpoints

Outcomes were collected twice more during follow-ups 2 weeks and 4 weeks after the experience, including the scores on the WEMWBS, modified version of the Tellegen Absorption Scale, QIDS, STAI-SF, and PDI.

Statistical Analysis

Logistic regression was used to assess the influence of the included variables on premature study termination, using the binary outcome attrition or completion as the dependent variable, where a participant was considered a completer if they had responded successfully to the key endpoint 4 weeks after the psychedelic experience. Owing to its nonreliance on the normal distribution of predictor variables, logistic regression is well-suited for the analysis of skewed data that have limited variability. Four models were fitted: 2 models containing baseline variables (1 including demographics, the other psychological measures), 1 containing postacute measures of psychedelic effects, and 1 with changes on relevant outcome variables between baseline and the 2-week endpoint. For those categorical and ordinal variables included in the demographic logit, we chose "Male" as reference category of gender owing to its greater size (n=485) compared to other levels of the variable, and the educational level "secondary education" owing to both its association with premature study termination known from the literature and its relevance in terms of sample size (n=276). Owing to a significant correlation (ρ=0.18; P<.001) and to reduce multicollinearity issues, it was decided to include only the educational level but not employment status as a predictor of attrition in the demographics logit. We also performed logistic regression analysis to investigate whether differences in psychedelic-induced changes on relevant outcome variables account for the likelihood of attrition. To compute change scores, individual baseline scores were subtracted from scores at the 2-week endpoint for each variable of interest (ie, WEMWBS, QIDS, STAI-SF, and PDI). Subsequently, a single logit was fitted, which included both change scores and absolute baseline scores, to account for the effect baseline variables. All statistical analyses were conducted using R (version 3.6.1, 2019-07-05; The R Foundation).

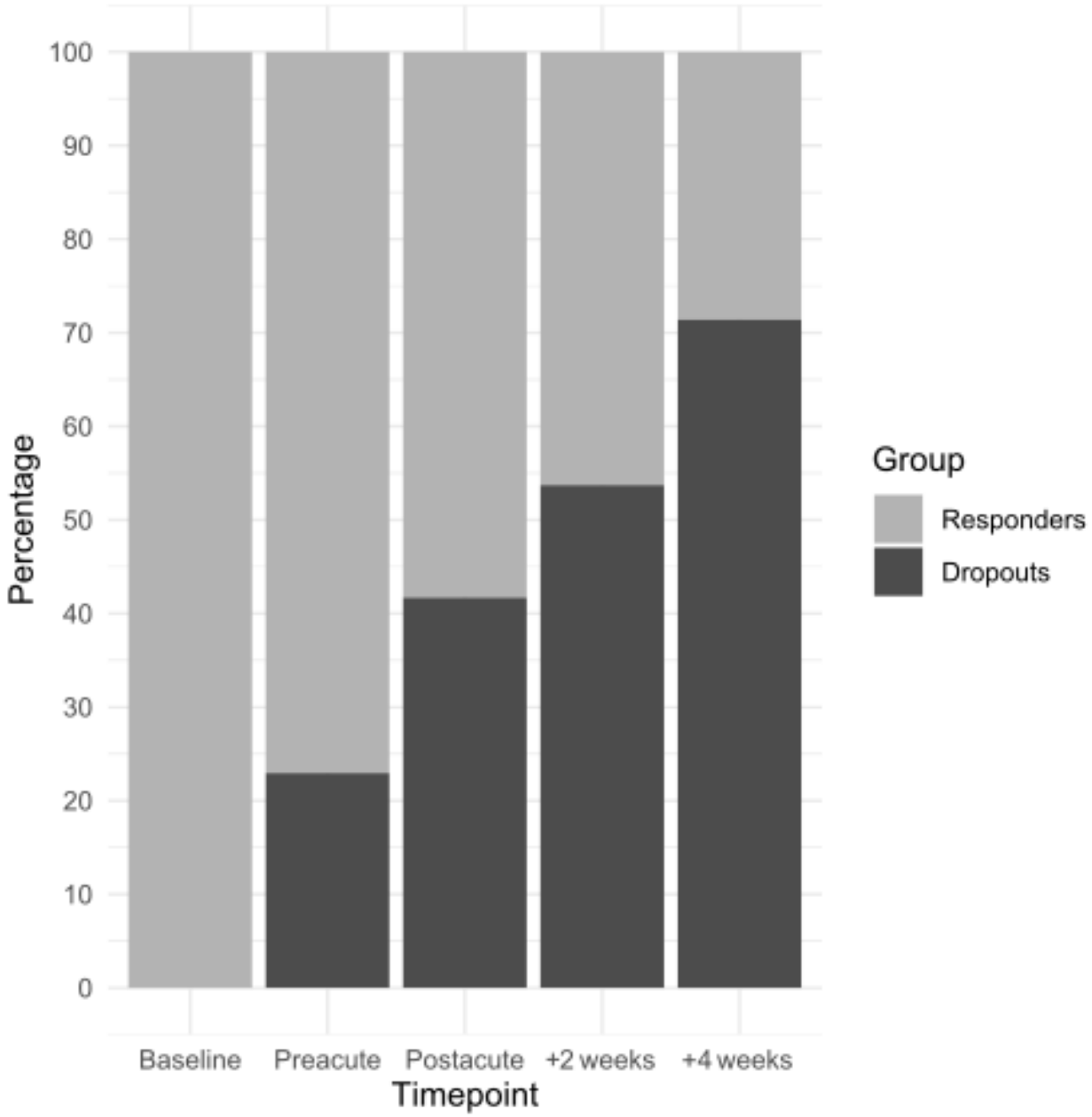
Results

Attrition Rates

In total, 564, 535, 379, 315, and 212 participants were sampled at survey timepoints 1, 2, 3, 4, and 5, respectively. Since some participants did not complete the baseline survey but responded at a subsequent timepoint, the overall number of participants was >654, (n=741). Of this total of 741 participants responding at any study timepoint, 529 (71.4%) subsequently dropped out at or prior to the final survey at 4 weeks post psychedelic experience. In particular, 170 (22.9%) participants stopped responding 1 day prior to the planned psychedelic experience (preacute timepoint), 139 (18.8%) stopped 1 day post experience (postacute timepoint), 89 (12%) stopped at the 2-week endpoint, and 131 (17.7%) stopped at the second follow-up and key endpoint 4 weeks post experience (Figure 1).

XSL•FO
RenderX

Figure 1. Cumulative frequency of participants dropping out (dark grey) and continuing to respond (light grey) to the key endpoint 4 weeks after a psychedelic experience among the sample sizes at each timepoint.



The distribution of dropouts according to timepoint was significantly different (χ^2_4=56.89; *P*<.001); more noncompleters dropped out prior to rather than after the postacute timepoint: 309 (58.4%) vs 220 (41.6%), respectively (χ^2_1=29.28; *P*<.001).

In total, 212 (28.6%) participants completed the second follow-up 4 weeks post experience and were thus classified as completers. Table 1 lists the demographics of participants at each timepoint of interest in this study.

XSL•FO
RenderX

Table 1. Demographics of participants at each time point.

Variables	Total	Dropouts	Completers
Baseline			
Number of participants	654	469	185
Age (years), mean (SD)	28.9 (10.4)	27.8 (9.9)	31.5 (11.5)
Gender, n (%)			
Male	485 (74.2)	359 (76.5)	126 (68.1)
Female	165 (25.2)	107 (22.8)	58 (31.4)
Other	4 (0.6)	3 (0.6)	1 (0.5)
Educational level, n (%)			
Primary education	53 (8.1)	34 (7.2)	19 (10.3)
Secondary education	276 (42.2)	223 (47.6)	53 (28.7)
University degree	325 (49.7)	212 (45.2)	113 (61.1)
Employment status, n (%)			
Unemployed	53 (8.1)	34 (7.3)	19 (10.3)
Student	256 (39.1)	198 (42.2)	58 (31.3)
Employed	335 (51.2)	232 (49.5)	103 (55.7)
Retired	10 (1.5)	5 (1.1)	5 (2.7)
Postacute timepoint (+1 day)			
Number of participants	379	192	187
Age (years), mean (SD)	30.6 (11.0)	29.8 (10.5)	31.5 (11.5)
Gender, n (%)			
Male	252 (66.5)	138 (71.9)	114 (61)
Female	97 (25.6)	42 (21.9)	55 (29.4)
Other	2 (0.5)	1 (0.5)	1 (0.5)
N/A[a]	28 (7.4)	11 (5.7)	17 (9.1)
Educational level, n (%)			
Primary education	30 (7.9)	12 (6.3)	18 (9.6)
Secondary education	121 (31.9)	76 (39.6)	45 (24.1)
University degree	200 (52.8)	93 (48.4)	107 (57.2)
N/A	28 (7.4)	11 (5.7)	17 (9.1)
Employment status, n (%)			
Unemployed	33 (8.7)	15 (7.8)	18 (9.6)
Student	124 (32.7)	71 (37.0)	53 (28.3)
Employed	188 (49.6)	94 (49)	94 (50.3)
Retired	6 (1.6)	1 (0.5)	5 (21.7)
N/A	28 (7.4)	11 (5.7)	17 (9.1)
First endpoint (+2 weeks), n (%)			
Number of participants	315	131	184
Age (years), mean (SD)	31.2 (11.2)	30.5 (10.7)	31.7 (11.6)
Gender, n (%)			
Male	196 (62.2)	82 (62.6)	114 (62.0)
Female	82 (26.0)	32 (24.4)	50 (27.2)
Other	1 (0.0)	0 (0.0)	1 (0.5)

XSL•FO
RenderX

Variables	Total	Dropouts	Completers
N/A	36 (11.4)	17 (13.0)	19 (10.3)
Educational level, n (%)			
Primary education	23 (7.3)	4 (3.1)	19 (10.3)
Secondary education	87 (27.6)	43 (32.6)	44 (23.9)
University degree	169 (53.7)	67 (51.1)	102 (55.4)
N/A	36 (11.4)	17 (13)	19 (10.3)
Employment status, n (%)			
Unemployed	27 (8.6)	11 (8.4)	16 (8.7)
Student	86 (27.3)	38 (29.0)	48 (26.1)
Employed	159 (50.5)	63 (48.1)	96 (52.2)
Retired	7 (2.2)	2 (1.5)	5 (6.7)
N/A	36 (11.4)	17 (13)	19 (10.3)

[a]Data at later timepoints were as not available for participants who had not completed the baseline questionnaire.

Logistic Regression

Table 2 presents the results of logistic regression analysis. Age significantly predicted attrition (β=–0.024; *P*=.007); specifically, older age was associated with a reduced probability of dropping out from the study. Participants with a university degree were less likely to drop out than those with a secondary educational level (β=–0.574; *P*=.005), and participants with primary education were also less likely to drop out than those with secondary education (β=–0.876; *P*=.008). Personality trait extraversion significantly predicted attrition (β=0.082; *P*=.012): higher scores on extraversion were associated with an increased probability of dropping out from the study. Personality trait conscientiousness also significantly predicted attrition (β=–0.079; *P*=.024) in the opposite direction: higher scores on conscientiousness were associated with a reduced probability of dropping out. All other assessed variables included in logistic regression analyses were nonsignificant in predicting attrition, including any measures of the acute psychedelic state and measures potentially related to adverse events, such as suicidality, delusional thinking, physical side effects, or challenging experiences. Similarly, no psychedelic-induced changes on outcome variables significant predicted study attrition.

XSL•FO
RenderX

Table 2. Results of logistic regression models predicting noncompletion of the study key endpoint.

Variables	β[a]	SE	z value	*P* value
Baseline demographics				
Intercept	1.633	0.509	3.209	.001
Age	–0.024	0.009	–2.684	*.007* [b]
Gender (reference: male)				
Female	–0.296	0.202	–1.460	.14
Other	–0.181	1.172	–0.155	.88
Educational level (reference: secondary education)				
University degree	–0.574	0.206	2.786	*.005*
Primary education	–0.876	0.329	–2.660	*.008*
Psychedelic advocacy	0.031	0.027	1.124	.26
Psychological variables				
Intercept	2.355	1.293	1.822	.07
Depression (QIDS[c])	0.018	0.031	0.598	.55
Well-being (WEMWBS[d])	–0.011	0.017	–0.618	.54
Personality (TIPI)				
Extraversion	0.082	0.037	2.518	*.01*
Conscientiousness	–0.079	0.035	–2.251	*.02*
Agreeableness	–0.018	0.041	–0.446	.66
Emotional stability	0.051	0.043	1.189	.23
Openness to experience	–0.044	0.052	–0.848	.40
Trait anxiety (STAI-SF[e])	–0.016	0.012	–1.377	.17
Social connectedness (SCS[f])	–0.007	0.012	–0.646	.52
Trait absorption (MODTAS[g])	–0.000	0.007	–0.080	.94
Suicidal ideation (SIDAS[h])	0.013	0.016	0.831	.41
Delusional thinking (PDI[i])	0.035	0.029	1.196	.24
Postacute timepoint (+1 day)				
Intercept	–0.257	0.317	–0.809	.42
Visual effects	0.002	0.002	1.078	.28
Challenging experience (CEQ[j])	0.001	0.007	–0.117	.91
Ego dissolution (EDI[k])	–0.010	0.006	–1.535	.13
Mystical experience (MEQ[l])	0.012	0.009	1.414	.16
Emotional breakthrough (EBI[m])	–0.007	0.005	–1.409	.16
Physical side-effects (PSI[n])	–0.010	0.057	–0.181	.86
Intercept for baseline–2-week changes	–1.590	2.312	–0.688	.49
Change scores (baseline to +2 weeks)				
Depression (QIDS)	–0.014	0.062	–0.221	.83
Psychological well-being (WEMWBS)	0.012	0.026	0.457	.65
Trait anxiety (STAI-SF)	0.027	0.024	1.137	.26
Social connectedness (SCS)	0.026	0.019	1.413	.16

XSL•FO
RenderX

Variables	β[a]	SE	*z* value	*P* value
Delusional thinking (PDI)	0.034	0.060	0.565	.57
Baseline control variables				
Depression (QIDS)	0.022	0.056	0.411	.68
Psychological well-being (WEMWBS)	0.001	0.029	0.028	.98
Trait anxiety (STAI-SF)	–0.014	0.021	–0.667	.51
Social connectedness (SCS)	0.019	0.018	1.019	.31
Delusional thinking (PDI)	0.044	0.039	1.144	.25

[a]β: estimated regression coefficient.
[b]Italicized values indicate significance levels of $P<.05$.
[c]QIDS: Quick Inventory of Depressive Symptomatology.
[d]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.
[e]STAI-SF: Spielberger State-Trait Anxiety Inventory.
[f]SCS: Social Connectedness Scale.
[g]MODTAS: modified version of the Tellegen Absorption Scale.
[h]SIDAS: Suicidal Ideation Attributes Scale.
[i]PDI: Peters' Delusional Inventory.
[j]CEQ: Challenging Experience Questionnaire.
[k]EDI: Ego Dissolution Inventory.
[l]MEQ: Mystical Experience Questionnaire.
[m]EBI: Emotional Breakthrough Inventory.
[n]PSI: Physical Symptoms Inventory.

Discussion

Principal Findings

This study aimed to identify variables accounting for attrition in prospective web-based studies on naturalistic psychedelic use. The overall attrition rate, which increased as a function of time, was high at 71.4% (n=529 of 741 participants). Contrary to prior hypotheses, neither the intensity of challenging experiences nor the advocacy of psychedelic use measured at baseline significantly predicted study completion. Rather, demographic variables including age and education, as well as personality traits including conscientiousness and extraversion, affected the likelihood of study noncompletion.

Specifically, logistic regression analyses revealed that young age, a low educational level, and the big 5 personality traits (high) extraversion and (low) conscientiousness were predictors of study attrition. The finding of age is in line with a number of previous studies [16-19] and may act synergistically with low conscientiousness to increase the likelihood of dropout. Indeed, several cross-sectional and longitudinal studies have shown that conscientiousness tends to increase with age [57-60], and that these increases often occur only during adulthood, which is relevant to the present study population, which had an average age of 28.9 (SD 10.4) years. The demographic logit also showed that participants with secondary education were more likely to drop out than those with university degrees and those with primary education. This may in part also be related to conscientiousness, which is known to be associated with educational expectations [61]. Highly conscientious young people also perform better academically and gain more advanced educational qualifications [62]. The personality trait conscientiousness is defined as "the propensity to follow socially prescribed norms and rules, to be goal-directed, planful, able to delay gratification, and to control impulses" [63]; thus, a conscientious person may indeed be more likely to commit to any obligation that he/she undertakes, be it loyalty to one's partner, paying taxes, or completion of a survey pertaining to psychedelic drug use. The finding that extraversion was a significant predictor of study attrition is also in line with previous studies reporting that extraversion predicts premature termination in longitudinal studies [34,64,65].

The absence of influence from any of the psychedelic-specific predictors is an important finding. Several previous studies [9-11,41] have indicated the quality of the acute psychedelic state to be a reliable predictor of longer-term psychological changes following psychedelic use. Recognizing the importance of acute subjective drug effects has been a key consideration informing the renewed interest in the therapeutic value of psychedelic compounds [41,66]. In the present study, neither the quality of the acute experience nor psychedelic advocacy or psychedelic-induced long-term psychological changes predicted study attrition. Given the accumulating number of studies reporting improvements in mental health outcomes after naturalistic psychedelic use [9,67-73], as well as impactful clinical trials involving psychedelic interventions [3,74-81], it can be considered reassuring that none of the established mediators of positive outcomes, nor outcomes themselves, seem to bias study attrition in longitudinal studies on psychedelics. Although it is, by definition, impossible to address with certainty how the noncompleters in this study sample faired in terms of their postpsychedelic mental health, the absence of a relationship between attrition and biased perspectives toward psychedelics

XSL•FO
RenderX

at baseline, the nature of the participants' acute experiences, or the reported beneficial effects (ie, psychological changes), partly ameliorates previous concerns regarding attrition bias in observational studies on psychedelics [9,10,39].

From among 529 dropouts, 309 (58.4%) stopped responding prior to the postacute timepoint, and approximately one-third did so even prior to the preacute timepoint, which could potentially be explained by impaired accessibility to and reduced desire to access the survey directly prior to and after the psychedelic experience or by the postponement or nonperformance of the experience. Although surveys could be completed through the mobile phone, it is conceivable that a portion of participants who consumed the substance under field conditions (eg, at a festival) stopped responding after the baseline assessment owing to pragmatic reasons of limited access as well as decayed motivation. As highlighted by others [29,30,82], predictors of attrition often differ for early vs late dropouts, and such differences may be important factors that allow for targeted interventions to reduce attrition; for instance, by activating reminders on mobile devices in the early phases of the prospective study. The effects of the substance use environment, associated intentions, and other contextual factors of psychedelic use on attrition should be targeted in future studies.

Limitations

Together, the observed effects suggest that the principal reasons for study attrition in observational studies on psychedelics are largely similar to those evidenced by other longitudinal studies that assessed phenomena independent of and unrelated to psychedelic use, with demographics and personality traits being the most predictive. Nevertheless, the present negative results with regard to potentially problematic systematic biases should only be considered preliminary, considering the limitations associated with the prevailing study design. Most significantly, the absence of data for those individuals who dropped out before completing the relevant surveys implies that both adverse experiences may have still occurred but were merely undetected. In future studies, with prior consent, this issue could be tackled by automatically sending out very brief surveys (eg, single-item surveys) to nonresponders, to investigate their reason for nonresponse. Similarly, rare cases of extreme negative reaction driving subsequent dropout may have been missed owing to both dropout and group averaging. Future studies focusing on such negative outliers may create pre-emptive value, given the disproportionate attention that can be attracted by such cases and the damaging impact this can have on broader studies and clinical development efforts. Further limitations of our study include a self-selection recruitment bias, which, as discussed by Haijen [9], reflects in a predominantly young, male, highly educated sample displaying strong psychedelic advocacy. On a more general level, the comparability of attrition studies across scientific domains and disciplines is a nontrivial problem. For example, outcome measures will not be consistent across studies, with the exception of simple demographic factors such as age. As revealed by independent studies [83-85], those on the predictors of attrition are often inconclusive, inconsistent, lacking in generalizability, and vulnerable to design-related limitations including a lack of standardization in definitions of attrition itself. Nevertheless, some evidence is convergent and our findings did generally converge consistently.

Conclusions

This study sought to identify factors accounting for the high attrition rates in a prospective study on naturalistic psychedelic use. Consistent with findings from other scientific disciplines, the strongest predictors of study attrition were observed among variables including age, educational level, and personality traits. In contrast, psychedelic-specific factors were found to be poor predictors of attrition. Methods for reducing attrition, which have been validated through other fields, such as text messaging [86], gamification [87,88], monetary incentives [89,90], or the creation of web-based participant communities [91], are thus likely to be applicable also to observational studies on psychedelics, which should be investigated in future studies. While not without prevailing limitations, for which there is significant scope for improvement, these findings somewhat support the reliability and validity of large-scale prospective web-based data collection as a methodology for studying the predictors and processes of changes related to psychedelic use.

Acknowledgments

RCH and his group are funded by supporters of the Centre for Psychedelic Research. RCH and HK are supported by the Ad Astra Chandaria Foundation, the Nikean Foundation, Tim Ferriss, Anton Bilton, the Tamas family, and the Alex Mosley Trust. We are grateful to Kenneth Jonck and Nikolai Lassen from the PsychedelicSurvey platform [42].

Conflicts of Interest

None declared.

References

1. Nestler S, Thielsch M, Vasilev E, Back MD. Will They Stay or Will They Go? Personality Predictors of Dropout in an Online Study. Int J Internet Sci 2015;10(1):37-48 [FREE Full text]
2. Froese T, Guzmán G, Guzmán-Dávalos L. On the Origin of the Genus Psilocybe and Its Potential Ritual Use in Ancient Africa and Europe. Econ Bot 2016 May 10;70(2):103-114. [doi: 10.1007/s12231-016-9342-2]

XSL•FO
RenderX

3. Carhart-Harris RL, Bolstridge M, Rucker J, Day CMJ, Erritzoe D, Kaelen M, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. Lancet Psychiatry 2016 Jul;3(7):619-627 [FREE Full text] [doi: 10.1016/S2215-0366(16)30065-7] [Medline: 27210031]
4. Sessa B. The 21st century psychedelic renaissance: heroic steps forward on the back of an elephant. Psychopharmacology (Berl) 2018 Feb;235(2):551-560. [doi: 10.1007/s00213-017-4713-7] [Medline: 28831571]
5. Nutt D, Erritzoe D, Carhart-Harris R. Psychedelic Psychiatry's Brave New World. Cell 2020 Apr 02;181(1):24-28 [FREE Full text] [doi: 10.1016/j.cell.2020.03.020] [Medline: 32243793]
6. Elsey JW. Psychedelic drug use in healthy individuals: A review of benefits, costs, and implications for drug policy. Drug Sci Policy Law 2017 Jul 26;3:205032451772323. [doi: 10.1177/2050324517723232]
7. Yockey RA, Vidourek RA, King KA. Trends in LSD use among US adults: 2015-2018. Drug Alcohol Depend 2020 Jul 01;212:108071. [doi: 10.1016/j.drugalcdep.2020.108071] [Medline: 32450479]
8. Krebs TS, Johansen P. Over 30 million psychedelic users in the United States. F1000Res 2013;2:98 [FREE Full text] [doi: 10.12688/f1000research.2-98.v1] [Medline: 24627778]
9. Haijen ECHM, Kaelen M, Roseman L, Timmermann C, Kettner H, Russ S, et al. Predicting Responses to Psychedelics: A Prospective Study. Front Pharmacol 2018;9:897 [FREE Full text] [doi: 10.3389/fphar.2018.00897] [Medline: 30450045]
10. Kettner H, Gandy S, Haijen ECHM, Carhart-Harris RL. From Egoism to Ecoism: Psychedelics Increase Nature Relatedness in a State-Mediated and Context-Dependent Manner. Int J Environ Res Public Health 2019 Dec 16;16(24):5147 [FREE Full text] [doi: 10.3390/ijerph16245147] [Medline: 31888300]
11. Roseman L, Haijen E, Idialu-Ikato K, Kaelen M, Watts R, Carhart-Harris R. Emotional breakthrough and psychedelics: Validation of the Emotional Breakthrough Inventory. J Psychopharmacol 2019 Sep;33(9):1076-1087. [doi: 10.1177/0269881119855974] [Medline: 31294673]
12. Lee S. An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey. Soc Sci Comput Rev 2016 Aug 19;24(4):460-475. [doi: 10.1177/0894439306288085]
13. Meade AW, Craig SB. Identifying careless responses in survey data. Psychol Methods 2012 Sep;17(3):437-455. [doi: 10.1037/a0028085] [Medline: 22506584]
14. Ward M, Meade AW, Allred CM, Pappalardo G, Stoughton JW. Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. Comput Hum Behav 2017 Nov;76:417-430. [doi: 10.1016/j.chb.2017.06.032]
15. Zhou H, Fishbach A. The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. J Pers Soc Psychol 2016 Oct;111(4):493-504. [doi: 10.1037/pspa0000056] [Medline: 27295328]
16. Honas JJ, Early JL, Frederickson DD, O'Brien MS. Predictors of attrition in a large clinic-based weight-loss program. Obes Res 2003 Jul;11(7):888-894 [FREE Full text] [doi: 10.1038/oby.2003.122] [Medline: 12855759]
17. Fabricatore AN, Wadden TA, Moore RH, Butryn ML, Heymsfield SB, Nguyen AM. Predictors of attrition and weight loss success: Results from a randomized controlled trial. Behav Res Ther 2009 Aug;47(8):685-691 [FREE Full text] [doi: 10.1016/j.brat.2009.05.004] [Medline: 19497559]
18. Gill RS, Karmali S, Hadi G, Al-Adra DP, Shi X, Birch DW. Predictors of attrition in a multidisciplinary adult weight management clinic. Can J Surg 2012 Aug;55(4):239-243 [FREE Full text] [doi: 10.1503/cjs.035710] [Medline: 22617538]
19. Jiandani D, Wharton S, Rotondi MA, Ardern CI, Kuk JL. Predictors of early attrition and successful weight loss in patients attending an obesity management program. BMC Obes 2016;3:14 [FREE Full text] [doi: 10.1186/s40608-016-0098-0] [Medline: 26966544]
20. Joyce AS, Piper WE, Ogrodniczuk JS. Therapeutic alliance and cohesion variables as predictors of outcome in short-term group psychotherapy. Int J Group Psychother 2007 Jul;57(3):269-296. [doi: 10.1521/ijgp.2007.57.3.269] [Medline: 17661544]
21. de Graaf R, Bijl RV, Smit F, Ravelli A, Vollebergh WA. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Am J Epidemiol 2000 Dec 01;152(11):1039-1047. [doi: 10.1093/aje/152.11.1039] [Medline: 11117613]
22. Agosti V, Nunes E, Ocepeck-Welikson K. Patient factors related to early attrition from an outpatient cocaine research clinic. Am J Drug Alcohol Abuse 1996 Feb;22(1):29-39. [doi: 10.3109/00952999609001643] [Medline: 8651143]
23. Daly J, Pelowski S. Predictors of dropout among men who batter: a review of studies with implications for research and practice. Violence Vict 2000;15(2):137-160. [Medline: 11108498]
24. Olver M, Stockdale KC, Wormith JS. A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. J Consult Clin Psychol 2011 Feb;79(1):6-21. [doi: 10.1037/a0022200] [Medline: 21261430]
25. Boekhout J, Peels DA, Berendsen BAJ, Bolman C, Lechner L. A Web-Based and Print-Delivered Computer-Tailored Physical Activity Intervention for Older Adults: Pretest-Posttest Intervention Study Comparing Delivery Mode Preference and Attrition. J Med Internet Res 2019 Aug 28;21(8):e13416 [FREE Full text] [doi: 10.2196/13416] [Medline: 31464186]
26. McDonald B, Haardoerfer R, Windle M, Goodman M, Berg C. Implications of Attrition in a Longitudinal Web-Based Survey: An Examination of College Students Participating in a Tobacco Use Study. JMIR Public Health Surveill 2017 Oct 16;3(4):e73 [FREE Full text] [doi: 10.2196/publichealth.7424] [Medline: 29038092]
27. Claus R, Kindleberger LR, Dugan MC. Predictors of attrition in a longitudinal study of substance abusers. J Psychoactive Drugs 2002;34(1):69-74. [doi: 10.1080/02791072.2002.10399938] [Medline: 12003115]

XSL•FO
RenderX

28. Wheeler E, Prettyman A, Lenhard MJ, Tran K. Adherence to outpatient program postoperative appointments after bariatric surgery. Surg Obes Relat Dis 2008;4(4):515-520. [doi: 10.1016/j.soard.2008.01.013] [Medline: 18586576]
29. Pekarik G, Blodgett C, Evans RG, Wierzbicki M. Variables related to continuance in a behavioral weight loss program. Addict Behav 1984;9(4):413-416. [doi: 10.1016/0306-4603(84)90044-3] [Medline: 6532151]
30. Clark MM, Niaura R, King TK, Pera V. Depression, smoking, activity level, and health status: pretreatment predictors of attrition in obesity treatment. Addict Behav 1996;21(4):509-513. [doi: 10.1016/0306-4603(95)00081-x] [Medline: 8830908]
31. Lamers F, Hoogendoorn AW, Smit JH, van Dyck R, Zitman FG, Nolen WA, et al. Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). Compr Psychiatry 2012 Jan;53(1):63-70 [FREE Full text] [doi: 10.1016/j.comppsych.2011.01.011] [Medline: 21397218]
32. Yass-Reed EM, Barry NJ, Dacey CM. Examination of pretreatment predictors of attrition in a VLCD and behavior therapy weight-loss program. Addict Behav 1993;18(4):431-435. [doi: 10.1016/0306-4603(93)90060-m] [Medline: 8213297]
33. Tsushima WT, Stoddard VM, Tsushima VG, Daly J. Characteristics of treatment drop-outs among two samples of chronic headache patients. J Clin Psychol 1991 Mar;47(2):199-205. [doi: 10.1002/1097-4679(199103)47:2<199::aid-jclp2270470204>3.0.co;2-h] [Medline: 2030124]
34. Hansson I, Berg AI, Thorvaldsson V. Can personality predict longitudinal study attrition? Evidence from a population-based sample of older adults. J Res Pers 2018 Dec;77:133-136. [doi: 10.1016/j.jrp.2018.10.002]
35. Topp BW, Kardash CA. Personality, achievement, and attrition: Validation in a multiple-jurisdiction police academy. Journal of Police Science & Administration 1986;14(3):234-241 [FREE Full text]
36. Meade AW, Pappalardo G. Predicting careless responses and attrition in survey data with personality. 2013 May Presented at: 28th Annual Meeting of the Society for Industrial and Organizational Psychology; 2013; Houston, TX.
37. Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. Arch Gen Psychiatry 1993 Dec;50(12):975-990. [doi: 10.1001/archpsyc.1993.01820240059008] [Medline: 8250684]
38. Vroling M, Wiersma FE, Lammers MW, Noorthoorn EO. Predicting Dropout from Intensive Outpatient Cognitive Behavioural Therapy for Binge Eating Disorder Using Pre-treatment Characteristics: A Naturalistic Study. Eur Eat Disord Rev 2016 Nov;24(6):494-502. [doi: 10.1002/erv.2474] [Medline: 27594180]
39. Kiefer D. Increasing Connection to Nature: A Novel Approach. Integr Med Alert 2020;24(2) [FREE Full text]
40. Kaertner L, Steinborn MB, Kettner H, Spriggs MJ, Roseman L, Buchborn T, et al. Positive expectations predict improved mental-health outcomes linked to psychedelic microdosing. Sci Rep 2021 Jan 21;11(1):1941 [FREE Full text] [doi: 10.1038/s41598-021-81446-7] [Medline: 33479342]
41. Roseman L, Nutt DJ, Carhart-Harris RL. Quality of Acute Psychedelic Experience Predicts Therapeutic Efficacy of Psilocybin for Treatment-Resistant Depression. Front Pharmacol 2017;8:974 [FREE Full text] [doi: 10.3389/fphar.2017.00974] [Medline: 29387009]
42. PsychedelicSurvey. URL: https://www.psychedelicsurvey.com/ [accessed 2021-02-14]
43. Alchemer. URL: https://app.alchemer.com/login/v1?redirected_from_legacy_domain=1 [accessed 2021-06-15]
44. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
45. Gosling S, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
46. Lee R, Robbins SB. Measuring belongingness: The Social Connectedness and the Social Assurance scales. J Couns Psychol 1995 Apr;42(2):232-241. [doi: 10.1037/0022-0167.42.2.232]
47. Jamieson GA. The modified tellegen absorption scale: A clearer window on the structure and meaning of absorption. Aust J Clin Exp Hypn 2005;33(2):119. [doi: 10.1037/t14465-000]
48. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
49. Rush A, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]
50. van Spijker BAJ, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The suicidal ideation attributes scale (SIDAS): Community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
51. Peters ER, Joseph SA, Garety PA. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). Schizophr Bull 1999;25(3):553-576. [doi: 10.1093/oxfordjournals.schbul.a033401] [Medline: 10478789]
52. Studerus E, Gamma A, Vollenweider FX. Psychometric evaluation of the altered states of consciousness rating scale (OAV). PLoS One 2010 Aug 31;5(8):e12412 [FREE Full text] [doi: 10.1371/journal.pone.0012412] [Medline: 20824211]

XSL•FO
RenderX

53. Barrett FS, Johnson MW, Griffiths RR. Validation of the revised Mystical Experience Questionnaire in experimental sessions with psilocybin. J Psychopharmacol 2015 Nov;29(11):1182-1190 [FREE Full text] [doi: 10.1177/0269881115609019] [Medline: 26442957]
54. Barrett FS, Bradstreet MP, Leoutsakos JS, Johnson MW, Griffiths RR. The Challenging Experience Questionnaire: Characterization of challenging experiences with psilocybin mushrooms. J Psychopharmacol 2016 Dec;30(12):1279-1295 [FREE Full text] [doi: 10.1177/0269881116678781] [Medline: 27856683]
55. Nour MM, Evans L, Nutt D, Carhart-Harris RL. Ego-Dissolution and Psychedelics: Validation of the Ego-Dissolution Inventory (EDI). Front Hum Neurosci 2016;10:269 [FREE Full text] [doi: 10.3389/fnhum.2016.00269] [Medline: 27378878]
56. Spector PE, Jex SM. Development of four self-report measures of job stressors and strain: Interpersonal Conflict at Work Scale, Organizational Constraints Scale, Quantitative Workload Inventory, and Physical Symptoms Inventory. J Occup Health Psychol 1998 Oct;3(4):356-367. [doi: 10.1037//1076-8998.3.4.356] [Medline: 9805281]
57. Helson R, Kwan VS, John OP, Jones C. The growing evidence for personality change in adulthood: Findings from research with personality inventories. J Res Pers 2002 Aug;36(4):287-306. [doi: 10.1016/S0092-6566(02)00010-7]
58. Helson R, Wink P. Personality change in women from the early 40s to the early 50s. Psychol Aging 1992 Mar;7(1):46-55. [doi: 10.1037//0882-7974.7.1.46] [Medline: 1558705]
59. Roberts BW, Walton KE, Viechtbauer W. Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. Psychol Bull 2006 Jan;132(1):1-25. [doi: 10.1037/0033-2909.132.1.1] [Medline: 16435954]
60. Srivastava S, John OP, Gosling SD, Potter J. Development of personality in early and middle adulthood: set like plaster or persistent change? J Pers Soc Psychol 2003 May;84(5):1041-1053. [doi: 10.1037/0022-3514.84.5.1041] [Medline: 12757147]
61. Cano M, Castillo LG, Davis MJ, López-Arenas A, Vaquero J, Thompson KV, et al. Dynamics of Acculturation, Enculturation, and Ethnic Identity: Influence of Psychocultural Constructs on Conscientiousness and Expectations of Higher Education Among Latinos. Int J Adv Counselling 2012 Apr 27;34(3):231-241. [doi: 10.1007/s10447-012-9153-9]
62. Almlund M, Duckworth AL, Heckman J, Kautz T. Personality Psychology and Economics. In: Hanushek EA, Machin S, Woessmann L, editors. Handbook of the Economics of Education (Volume 4). Amsterdam: Elsevier; 2011:1-181.
63. John OP, Srivastava S. The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin LA, John OP, editors. Handbook of personality: Theory and research. New York, NY: Guilford Press; 1999:102-138.
64. Lugtig P. Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. Sociol Methods Res 2014 Feb 06;43(4):699-723. [doi: 10.1177/0049124113520305]
65. Satherley N, Milojev P, Greaves LM, Huang Y, Osborne D, Bulbulia J, et al. Demographic and psychological predictors of panel attrition: evidence from the New Zealand attitudes and values study. PLoS One 2015;10(3):e0121950 [FREE Full text] [doi: 10.1371/journal.pone.0121950] [Medline: 25793746]
66. Carhart-Harris RL, Roseman L, Haijen E, Erritzoe D, Watts R, Branchi I, et al. Psychedelics and the essential importance of context. J Psychopharmacol 2018 Jul;32(7):725-731. [doi: 10.1177/0269881118754710] [Medline: 29446697]
67. Zeifman RJ, Wagner AC, Watts R, Kettner H, Mertens LJ, Carhart-Harris RL. Post-Psychedelic Reductions in Experiential Avoidance Are Associated With Decreases in Depression Severity and Suicidal Ideation. Front Psychiatry 2020;11:782 [FREE Full text] [doi: 10.3389/fpsyt.2020.00782] [Medline: 32903724]
68. Uthaug MV, Lancelotta R, van Oorsouw K, Kuypers KPC, Mason N, Rak J, et al. A single inhalation of vapor from dried toad secretion containing 5-methoxy-N,N-dimethyltryptamine (5-MeO-DMT) in a naturalistic setting is related to sustained enhancement of satisfaction with life, mindfulness-related capacities, and a decrement of psychopathological symptoms. Psychopharmacology (Berl) 2019 Sep;236(9):2653-2666 [FREE Full text] [doi: 10.1007/s00213-019-05236-w] [Medline: 30982127]
69. Uthaug MV, van Oorsouw K, Kuypers KPC, van Boxtel M, Broers NJ, Mason NL, et al. Sub-acute and long-term effects of ayahuasca on affect and cognitive thinking style and their association with ego dissolution. Psychopharmacology (Berl) 2018 Oct;235(10):2979-2989 [FREE Full text] [doi: 10.1007/s00213-018-4988-3] [Medline: 30105399]
70. Mason NL, Mischler E, Uthaug MV, Kuypers KPC. Sub-Acute Effects of Psilocybin on Empathy, Creative Thinking, and Subjective Well-Being. J Psychoactive Drugs 2019;51(2):123-134. [doi: 10.1080/02791072.2019.1580804] [Medline: 30905276]
71. Soler J, Elices M, Dominguez-Clavé E, Pascual JC, Feilding A, Navarro-Gil M, et al. Four Weekly Ayahuasca Sessions Lead to Increases in "Acceptance" Capacities: A Comparison Study With a Standard 8-Week Mindfulness Training Program. Front Pharmacol 2018;9:224 [FREE Full text] [doi: 10.3389/fphar.2018.00224] [Medline: 29615905]
72. Barbosa PCR, Cazorla IM, Giglio JS, Strassman R. A six-month prospective evaluation of personality traits, psychiatric symptoms and quality of life in ayahuasca-naïve subjects. J Psychoactive Drugs 2009 Sep;41(3):205-212. [doi: 10.1080/02791072.2009.10400530] [Medline: 19999673]
73. Spriggs MJ, Kettner H, Carhart-Harris RL. Positive effects of psychedelics on depression and wellbeing scores in individuals reporting an eating disorder. Eat Weight Disord 2021 May;26(4):1265-1270. [doi: 10.1007/s40519-020-01000-8] [Medline: 32895801]

XSL•FO
RenderX

74. Palhano-Fontes F, Barreto D, Onias H, Andrade KC, Novaes MM, Pessoa JA, et al. Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. Psychol Med 2019 Mar;49(4):655-663 [FREE Full text] [doi: 10.1017/S0033291718001356] [Medline: 29903051]
75. Carhart-Harris RL, Bolstridge M, Day CMJ, Rucker J, Watts R, Erritzoe DE, et al. Psilocybin with psychological support for treatment-resistant depression: six-month follow-up. Psychopharmacology (Berl) 2018 Feb;235(2):399-408 [FREE Full text] [doi: 10.1007/s00213-017-4771-x] [Medline: 29119217]
76. Grob CS, Danforth AL, Chopra GS, Hagerty M, McKay CR, Halberstadt AL, et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. Arch Gen Psychiatry 2011 Jan;68(1):71-78. [doi: 10.1001/archgenpsychiatry.2010.116] [Medline: 20819978]
77. Johnson MW, Garcia-Romeu A, Cosimano MP, Griffiths RR. Pilot study of the 5-HT2AR agonist psilocybin in the treatment of tobacco addiction. J Psychopharmacol 2014 Nov;28(11):983-992 [FREE Full text] [doi: 10.1177/0269881114548296] [Medline: 25213996]
78. Griffiths RR, Johnson MW, Carducci MA, Umbricht A, Richards WA, Richards BD, et al. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. J Psychopharmacol 2016 Dec;30(12):1181-1197 [FREE Full text] [doi: 10.1177/0269881116675513] [Medline: 27909165]
79. Anderson B, Danforth A, Daroff R, Stauffer C, Diller J, Woolley J. Psilocybin-assisted Group Therapy for Demoralization in Long-term AIDS Survivors. Biol Psychiatry 2019;85:S182 [FREE Full text]
80. Bogenschutz MP, Forcehimes AA, Pommy JA, Wilcox CE, Barbosa P, Strassman RJ. Psilocybin-assisted treatment for alcohol dependence: a proof-of-concept study. J Psychopharmacol 2015 Mar;29(3):289-299. [doi: 10.1177/0269881114565144] [Medline: 25586396]
81. Ross S, Bossis A, Guss J, Agin-Liebes G, Malone T, Cohen B, et al. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. J Psychopharmacol 2016 Dec;30(12):1165-1180 [FREE Full text] [doi: 10.1177/0269881116675512] [Medline: 27909164]
82. Bennett GA, Jones SE. Dropping out of treatment for obesity. J Psychosom Res 1986;30(5):567-573. [doi: 10.1016/0022-3999(86)90029-2] [Medline: 3772838]
83. Sayre SL, Schmitz JM, Stotts AL, Averill PM, Rhoades HM, Grabowski JJ. Determining predictors of attrition in an outpatient substance abuse program. Am J Drug Alcohol Abuse 2002;28(1):55-72. [doi: 10.1081/ada-120001281] [Medline: 11853135]
84. Warnick EM, Gonzalez A, Robin Weersing V, Scahill L, Woolston J. Defining dropout from youth psychotherapy: how definitions shape the prevalence and predictors of attrition. Child Adolesc Ment Health 2012 May;17(2):76-85. [doi: 10.1111/j.1475-3588.2011.00606.x] [Medline: 32847293]
85. Davis MJ, Addis ME. Predictors of attrition from behavioral medicine treatments. Ann Behav Med 1999;21(4):339-349. [doi: 10.1007/BF02895967] [Medline: 10721442]
86. Grutzmacher SK, Munger AL, Speirs KE, Vafai Y, Hilberg E, Braunscheidel Duru E, et al. Predicting Attrition in a Text-Based Nutrition Education Program: Survival Analysis of Text2BHealthy. JMIR Mhealth Uhealth 2019 Jan 21;7(1):e9967 [FREE Full text] [doi: 10.2196/mhealth.9967] [Medline: 30664489]
87. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and Adherence to Web-Based Mental Health Interventions: A Systematic Review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
88. Lumsden J, Skinner A, Coyle D, Lawrence N, Munafo M. Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques. J Med Internet Res 2017 Nov 22;19(11):e395 [FREE Full text] [doi: 10.2196/jmir.8473] [Medline: 29167090]
89. Khadjesari Z, Murray E, Kalaitzaki E, White IR, McCambridge J, Thompson SG, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011 Mar 02;13(1):e26 [FREE Full text] [doi: 10.2196/jmir.1523] [Medline: 21371988]
90. Becker R, Glauser D. Are Prepaid Monetary Incentives Sufficient for Reducing Panel Attrition and Optimizing the Response Rate? An Experiment in the Context of a Multi-Wave Panel with a Sequential Mixed-Mode Design. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique 2018 May 08;139(1):74-95. [doi: 10.1177/0759106318762456]
91. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]

Abbreviations

PDI: Peters' Delusional Inventory
QIDS: Quick Inventory of Depressive Symptomatology
SIDAS: Suicidal Ideation Attributes Scale
STAI-SF: Spielberger State-Trait Anxiety Inventory
WEMWBS: Warwick-Edinburgh Mental Well-being Scale

XSL•FO
RenderX

Edited by R Kukafka; submitted 24.11.20; peer-reviewed by S Baee, C Entwistle; comments to author 09.02.21; revised version received 19.02.21; accepted 04.05.21; published 28.07.21.

Please cite as:
Hübner S, Haijen E, Kaelen M, Carhart-Harris RL, Kettner H
Turn on, Tune in, and Drop out: Predictors of Attrition in a Prospective Observational Cohort Study on Psychedelic Use
J Med Internet Res 2021;23(7):e25973
URL: https://www.jmir.org/2021/7/e25973
doi:10.2196/25973
PMID:34319246

©Sebastian Hübner, Eline Haijen, Mendel Kaelen, Robin Lester Carhart-Harris, Hannes Kettner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 28.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Preadolescent Students' Engagement With an mHealth Intervention Fostering Social Comparison for Health Behavior Change: Crossover Experimental Study

Raoul Ceasar Yannic Nuijten[1], MSc; Pieter Van Gorp[1], PhD; Tom Borghouts[1], BSc; Pascale Le Blanc[1], PhD; Pauline Van den Berg[2], PhD; Astrid Kemperman[2], PhD; Ehsan Hadian[1], MSc; Monique Simons[3], PhD

[1]Department of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[2]Department of the Built Environment, Eindhoven University of Technology, Eindhoven, Netherlands

[3]Department of Social Sciences, Wageningen University and Research, Wageningen, Netherlands

Corresponding Author:
Raoul Ceasar Yannic Nuijten, MSc
Department of Industrial Engineering
Eindhoven University of Technology
Groene Loper 3
Eindhoven, 5612 AE
Netherlands
Phone: 31 0614906142
Email: r.c.y.nuijten@tue.nl

Abstract

Background: Contemporary mobile health (mHealth) interventions use various behavior change techniques to promote healthier lifestyles. Social comparison is one of the techniques that is consensually agreed to be effective in engaging the general population in mHealth interventions. However, it is unclear how this strategy can be best used to engage preadolescents. Nevertheless, this strategy has great potential for this target audience, as they are particularly developing their social skills.

Objective: This study aims to evaluate how social comparison drives preadolescents' engagement with an mHealth app.

Methods: We designed a 12-week crossover experiment in which we studied 3 approaches to implementing behavior change via social comparison. This study was hosted in a school environment to leverage naturally existing social structures among preadolescents. During the experiment, students and teachers used an mHealth tool that awarded points for performing healthy activities. Participants could read their aggregated scores on a leaderboard and compare their performance with others. In particular, these leaderboards were tweaked to implement 3 approaches of the social comparison technique. The first approach focused on intragroup comparison (ie, students and teachers competing against each other to obtain the most points), whereas the other two approaches focused on intergroup comparison (ie, classes of students and their mentoring teachers collaborating to compete against other classes). Additionally, in the third approach, the performance of teachers was highlighted to further increase students' engagement through teachers' natural exemplary function. To obtain our results, we used linear modeling techniques to analyze the dropout rates and engagement levels for the different approaches. In such analyses, we also considered individual participant traits.

Results: Our sample included 313 participants—290 students (92.7%) and 23 teachers (7.3%). It was found that student engagement levels dropped over time and declined during holidays. However, students seemed to monitor the intergroup competitions more closely than the intragroup competitions, as they, on average, checked the mHealth app more often when they were engaged in team-based comparisons. Students, on average, performed the most unique activities when they were engaged in the second intergroup setting, perhaps because their teachers were most active in this setting. Moreover, teachers seemed to play an important role in engaging their students, as their relationship with their students influenced the engagement of the students.

Conclusions: When using social comparison to engage preadolescents with an mHealth tool, an intergroup setting, rather than an intragroup competition, motivated them to engage with the app but did not necessarily motivate them to perform more activities. It seems that the number of unique activities that preadolescents perform depends on the activeness of a role model. Moreover, this effect is amplified by preadolescents' perceptions of closeness to that role model.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e21202) doi:10.2196/21202

KEYWORDS

mHealth; health promotion; social comparison; competitiveness; collaboration; gamification; preadolescents; high school students

Introduction

Research Case

Many people with chronic diseases and other health-related problems may benefit greatly from increased physical activity and improved dietary intake [1]. To support individuals adopting these healthier routines, behavior change interventions, including various behavior change techniques to encourage certain targeted behaviors, may be used [2]. The aim of this study is to strengthen the empirical evidence on the impact of one specific technique (ie, social comparison) for engaging a specific target group—in our case, preadolescents (ie, 10- to 13-year-olds).

We have chosen to target preadolescents because application of interventions at this stage in life will likely also impact health at a later stage, particularly because newly adopted lifestyle behaviors will track into adulthood [3]. By the time of preadolescence, the human brain is still developing, particularly strong in the areas of social skills and peer relationships [4]. Therefore, when designing an intervention for preadolescents, taking into account the social dynamics within that target group is of vital importance for intervention effectiveness.

At the same time, social dynamics have been used to foster health behavior changes. Social comparison is one of the behavior change techniques that is consensually agreed to be effective for the general population [2]. This intervention strategy stems from the idea that, by nature, people tend to self-evaluate by comparing themselves with others [5]. Comparing oneself is a universal process that all of us engage in regularly, although some more than others [6]. Similarly, competitive processes, as manifestations of social comparison [7], are prevalent in our societies [8]. As these phenomena are common in everyday life, it seems beneficial to evaluate whether these processes can be applied to promote health behavior change.

In mobile health (mHealth) interventions, the adoption of leaderboards offers an opportunity to implement social comparison and competitiveness as intervention strategies [9]. Leaderboards are a form of gamification, a set of motivational techniques that use game mechanics outside game contexts, to foster participation, engagement, and loyalty [10,11]. Leaderboards may be used to increase participant engagement and are widely adopted as one of the most popular gamification techniques [12]. In this study, we investigate how to design such leaderboards for the optimal engagement of preadolescents.

The degree to which individuals are engaged in a particular social comparative setting is determined by different situational and individual factors [8]. A key decision when designing such a setting is to determine whether comparative and competitive processes occur either intra- or intergroup [13]. In an intragroup setting, individuals compete against each other, whereas in an intergroup setting, groups of individuals collaborate to compete against other groups.

In this study, we evaluate the implementation of social comparison (ie, either fostering intra- or intergroup comparisons) that is most effective in promoting healthy routines in preadolescents. As social dynamics among preadolescents are likely articulated in their school environment and because the World Health Organization has put forward the key objective of "[generating] scientific evidence on effective Health Promoting School (HPS) interventions" [4], we have implemented our intervention at a high school. In this environment, educational levels and classes are the main social structures. Moreover, teachers have an exemplary function within a school environment and potentially serve as positive role models for students [14]. In our experiment, we analyzed the role of social dynamics between students and teachers in three study arms.

We hypothesized that an intergroup approach, which combines collaborative and competitive aspects, would be engaging for students and would encourage them to adopt new healthy routines. We based our hypothesis on the observation that an intergroup approach can potentially trigger processes of self-enhancement in children and processes of enhancement of others, whereas intragroup competition is likely to solely promote self-enhancement, potentially even at the expense of others [15,16]. This claim is supported by a recent review of competitive versus cooperative aspects in social exergames: cooperative play can "increase motivation, promote continued play, and increase prosocial behaviors," whereas competitive play mainly yields short-term, physiological arousal [17]. To empirically evaluate whether the potential spillover effects of intergroup competitions do indeed positively influence students' engagement levels in a health promotion campaign, we designed study arms that ranged from more intragroup to more intergroup focused. Before explaining these treatments in detail in the *Methods* section, the factors influencing the social comparative behavior are evaluated in depth.

Theoretical Background

By nature, people tend to self-evaluate by comparing themselves with others [5]. This study evaluates how this natural tendency may be leveraged to sustain the engagement of preadolescent students in a health promotion campaign. Comparing ourselves with others may occur in different directions; we may compare ourselves, based on a specific aspect, with others who are worse (ie, downward comparison) or better (ie, upward comparison). Both downward and upward comparisons may affect one's self-concept [7] and can foster competitive behavior [8]. Downward comparison will often enhance the self-concept of the comparator [18]. However, as downward comparison reveals that ones' status could decline if others catch up, the feeling of being threatened and discouragement might also be evoked [18]. Comparing ourselves with a superior other might cause negative feelings too, as the other performs better on certain attributes [19]. At the same time, upward comparison can lead to the assimilation of the characteristics of the superior other and

XSL•FO
RenderX

provide hope and inspiration, especially if the superior is a role model (eg, a student's teacher) [20].

The social comparison model of competition describes the factors that influence competitive behavior [8]. The model proposes that competitive processes are influenced by situational factors. For example, it was found that the number of competitors is best kept as low as possible: the lower the number of competitors, the more intense the competition (ie, the N effect [21,22]). Furthermore, it was found that incentives, such as tangible rewards, increase people's engagement in a competitive setting [8,23,24].

In addition, the social comparison model of competition proposes that competitive processes are influenced by individual factors. For example, it was found that, for a competition to be engaging, participants have to perceive the dimension of comparison as relevant to the self [25]. This effect is particularly amplified if competitors perceive their relationships as *close* [26,27]. Furthermore, it was found that when competitors, either as a group or as individuals, share similar characteristics (eg, race or education), the competition intensifies [7,28]. Similarly, personality traits are known to increase competitiveness. In particular, social comparison orientation [6], competitive dispositions [29], and individuals' orientation toward performance goals [30,31] seem to influence competitiveness. Moreover, the personality trait openness to experience (ie, as defined by the Big Five personality framework [32]) is a potential trait that can influence competitiveness: people who score lower on this trait and are therefore less independent and creative may be more competitive [33].

Finally, it has been previously demonstrated that especially intergroup competitions can enhance engagement in an activity [13] because it includes a mix of collaborative and competitive aspects. It was found that an intergroup competition can potentially trigger processes of self-enhancement in children and processes of enhancement of others, whereas intragroup competitions are likely to solely promote self-enhancement, potentially even at the expense of others [15,16]. These principles were not yet tested with preadolescents in schools.

This study aims to further such theoretical insights via a study design that contrasts inter- and intragroup competition in schools and that tests whether teachers as role models can increase engagement. Potential moderation of situational factors and personality traits is accounted for.

Methods

Recruitment

Participants were recruited among first-year, prevocational (ie, VMBO [Voorbereidend Middelbaar BeroepsOnderwijs]) students (ie, 11- to 13-year-olds) at a high school in the Netherlands in April 2019. The study was advertised as a health promotion campaign and conducted only after obtaining explicit written consent of the participants (ie, the students) and their parents or guardians. Explicit consent of the students was collected upon registration for the campaign. Explicit consent from their parents or guardians was collected via consent letters that they signed and returned to the school's administration. More operational procedures were also approved by the ethical committee of Eindhoven University (Archie experiment ID 920). The ethical review committee concluded that the potential benefits of this study outweighed its potential risks. However, besides the potential positive impact of social comparison on health behavior, it was acknowledged that the target group may have also experienced the negative effects of social comparison (eg, feeling threatened or discouraged [18]). Meanwhile, it was found that the school environment provided a sufficiently safe setting to host this experiment, especially because teachers were advised to check in weekly with their students on the impact of the campaign. As such, the more vulnerable students could have been identified and corrective actions could have been taken.

Intervention

To test our hypotheses, we have used the mHealth tool GameBus that is manufactured and maintained by Eindhoven University of Technology. GameBus was especially designed for health promotion and provides a highly configurable gamification engine that is used to sustain participants' engagement. According to the classification of gamification elements by Hamari et al [12], GameBus implements the gamification mechanisms *challenges*, *points*, *goals*, *progress*, *leaderboards*, and *rewards* and configures these mechanisms to test scientific hypotheses. The tool supports hosting multiple experimental designs on a single platform, ensuring that user experience remains similar across these different designs. At the same time, the platform enables researchers to gather rich data in a manner compliant with European privacy legislations. For this study, a dedicated Privacy Impact Assessment was approved by the Data Protection Officers of Eindhoven University of Technology and the high school.

GameBus includes a mobile app (ie, available via the web as well as on Android and iOS) for tracking healthy activities. The platform is designed to allow rewarding any healthy activity with points, from social, to mental, to physical activities. For this study, the rewarded activities have been defined in consultation with the school's management and a student council. Several cocreation sessions were held with the aim of defining activities that students were capable of performing, that they would enjoy doing, and that would benefit their health. The mHealth tool would then reward students for performing these activities, based on a selfie as proof of conducting the activity.

Users could compare their performance on a leaderboard that summed up the points per user. In addition, for the intergroup approaches, additional leaderboards showed per team (ie, per class) the average number of points across team members. The app also provided a newsfeed, which showed an entry for each team member that scored points. Such entries could be liked or commented upon in a manner similar to mainstream social media platforms such as Facebook and Instagram.

The overall goal of the intervention from the students' perspective was to obtain as many points as possible by adopting healthier routines. In particular, it was set by the school's management to focus on (1) increasing physical activity; (2) promoting healthy nutrition; (3) fostering sustainable relationships: friends, love, and intimacy; and (4) emphasizing the (potential) impact of stress, drugs, alcohol, and gaming.

XSL•FO
RenderX

From these focal areas, a list of prescribed activities was compiled in consultation with the school's management and a student council. The aim was to define activities that students were capable of performing, that they would enjoy doing, and that would benefit their health (eg, "Wrestle arms with someone of at least 40+" or "Peel an (unbroken) apple peel of at least 20 centimeters"), resulting in a list of 51 unique activities. These activities were distributed over the course of the campaign and renewed every *wave* (ie, every 2 weeks; a complete overview of prescribed activities per wave is provided in Multimedia Appendix 1). The entire campaign lasted 12 weeks (ie, 6 waves). Eventually, the first wave consisted of 12 unique activities, and in the other five waves, nine activities were prescribed. Each wave included a mix of the focal areas. Some activities were duplicated over multiple waves (a detailed overview is provided in Multimedia Appendix 1).

Study Design

Treatment Allocation

From a scientific perspective, the intervention included three different social comparative settings as treatments, to test whether an intergroup—rather than an intragroup—competition would be more effective in promoting healthy routines in preadolescent students. A crossover study design was adopted to ensure that all the participants were exposed twice to every treatment. We adopted a randomized block approach to randomly distribute the treatments to the participants. By order of the school's management, the three clusters that this study design required were defined based on educational level (Dutch prevocational education distinguishes three such levels). Participants received the treatments in 2-week periods (ie, in so-called waves). The entire campaign lasted for 12 weeks (ie, 6 waves); therefore, each participant received every treatment twice.

Our treatments effectively simulated three different implementations of the social comparison technique. One approach represented an intragroup competition, whereas the other two approaches represented intergroup competitions. In one of these intergroup competitions, an additional comparative element was introduced (ie, by explicitly highlighting the performance of teachers), as this manipulation was expected to increase students' engagement levels even further, because teachers potentially serve as role models for students and may, therefore, foster hope and inspiration in students during the competition, particularly if they perform (somewhat) better than their students [20]. Note that, although the treatments were different in nature, the overall objective was always the same from the participants' perspective—to collect as many points as possible by performing the healthy activities prescribed in each specific wave. The player, or team, with the greatest number of points at the end of a wave (ie, the absolute winner) was awarded a small gift (ie, either a medal or a stress ball). The following paragraphs describe the different treatments in detail. Table 1 displays how the treatments were distributed across the participants. The rows distinguish between the three treatment groups. In the columns, it can be read what treatment each treatment group is assigned in a given wave.

Table 1. Distribution of treatments over participants and waves.

Educational level	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Educational level A	SCS2[a]	SCS1[b]	SCS3[c]	SCS2	SCS1	SCS3
Educational level B	SCS3	SCS2	SCS1	SCS3	SCS2	SCS1
Educational level C	SCS1	SCS3	SCS2	SCS1	SCS3	SCS2

[a]SCS2: second social comparative setting.

[b]SCS1: first social comparative setting.

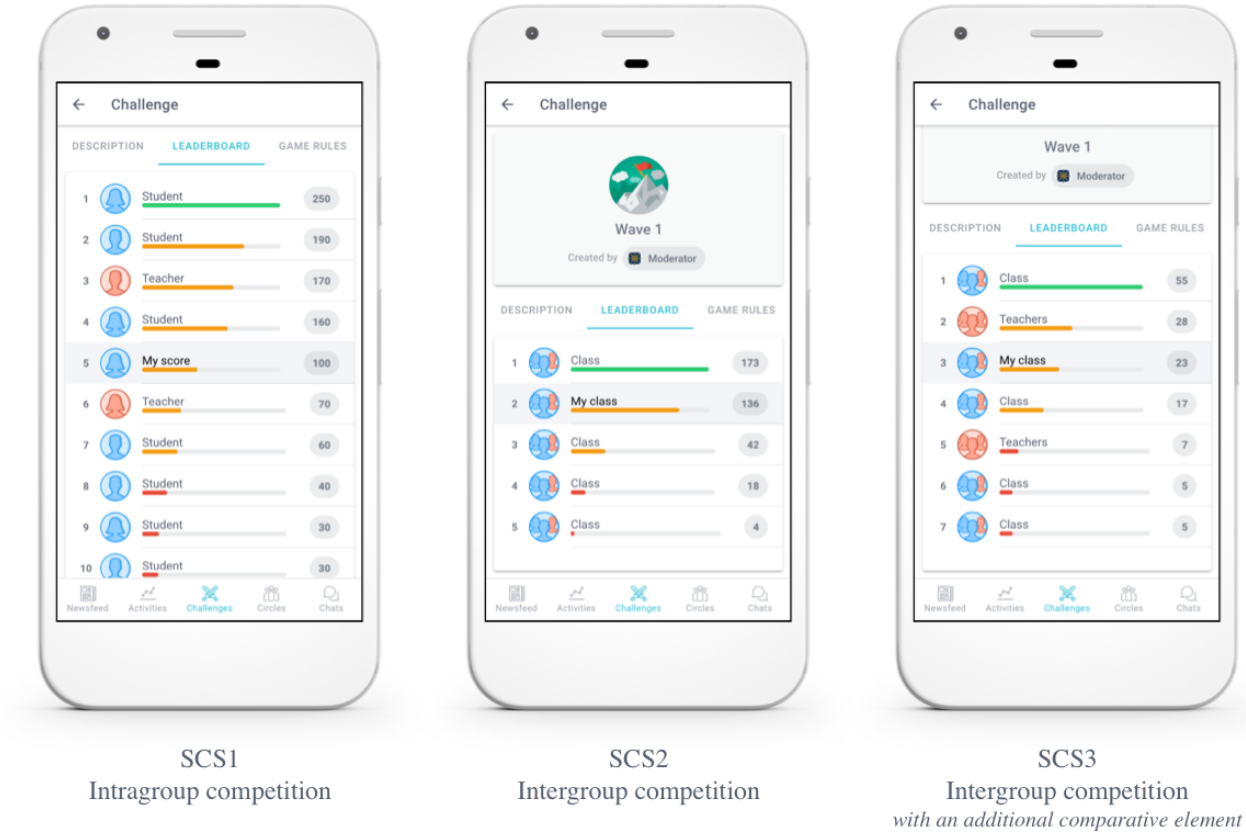
[c]SCS3: third social comparative setting.

First Social Comparative Setting: Intragroup Competition

In the first social comparative setting (SCS1), students and teachers of a treatment group (ie, educational level) competed individually. In this intragroup competition, players could read each other's performance from a leaderboard but could not see the actual activities, other than their own, that were performed to accumulate this score in the newsfeed of the mHealth tool (Figure 1). Only the absolute winner was awarded a small gift for this competition type.

XSL•FO
RenderX

Figure 1. Leaderboard view of different treatments (students' and class' names removed for the sake of confidentiality). SCS1: first social comparative setting; SCS2: second social comparative setting; SCS3: third social comparative setting.



SCS1
Intragroup competition

SCS2
Intergroup competition

SCS3
Intergroup competition
with an additional comparative element

Second Social Comparative Setting: Intergroup Competition

In the second social comparative setting (SCS2), students of a particular class and their mentoring teachers joined forces to compete against other classes within their treatment group (Figure 1). In this intergroup competition, players could read their class' performance (ie, the average number of points collected by the members of their class) on the leaderboard. In addition, they could read their own contribution to the score of their class relative to the contribution of their class members and mentors, but they could not read the individual contributions of students and mentors of other classes. Similarly, players could see the actual activities that others performed to accumulate their score if, and only if, that player was within their own class. At the end of the wave, the entire winning team was awarded a small gift.

Third Social Comparative Setting: Intergroup Competition With an Additional Dimension of Comparison

The third social comparative setting (SCS3) closely resembled the second treatment: this setting also featured an intergroup competition in which the entire winning team was awarded a small gift at the end of a wave. In SCS2 however, students could not transparently compare their performance with their teachers (other than their own mentoring teachers) because a teacher's score was concealed in the average score of the class they mentor and can, therefore, not be read by students from other classes. However, Lockwood and Kunda [20] argue that students can draw inspiration from the act of comparing themselves with their teachers, especially if their teachers (slightly) outperform them (ie, triggering upward comparison) [20]. Therefore, to allow students to compare themselves with their teachers, SCS3 included two extra teams that were composed of teachers only (Figure 1). These two teams included teachers who mentor students from the two other treatment groups (ie, the two other educational levels). Therefore, in SCS3, students would collaborate with their class members and mentoring teachers to compete against other groups of students and teachers.

Study Procedures

At the start of the campaign, a kick-off day was scheduled. Before this day, teachers were introduced to the mHealth app and study context. Subsequently, the teachers instructed their students on using the mobile app. Throughout the kick-off day, a dedicated support team was present to assist students with installing the mHealth app. In addition, several workshops were organized by the Public Health Services to introduce students to topics such as healthy nutrition, the dangers of smoking, alcohol abuse, and the use of drugs.

To keep the campaign under the attention of students for the entire 12-week period, teachers were instructed to discuss their students' progress in class once a week. It was particularly suggested to review in plenary class sessions the leaderboard and discuss the activities the students had performed over the

XSL•FO
RenderX

last week. Unfortunately, this review could not take place in the third and fourth weeks (ie, during the second wave) of the campaign owing to the spring break.

Measurements

In mHealth, engagement is most commonly captured via passive measures of app use [34]. Using the GameBus platform, engagement of participants was repeatedly measured as (1) the number of days a participant had visited the app, and (2) the number of unique activities a respondent performed. These variables complement each other because the former may be limited to passive engagement, whereas the latter requires active participation (ie, performing healthy activities). Note that our second outcome variable measures the number of unique activities a respondent performed rather than the total number of activities that were performed. The main reason is that we aim to encourage preadolescents to adopt a multitude of healthy routines, not just repeat a single routine (ie, quality over quantity). However, our results and conclusions did not differ when analyzing the total number of activities that respondents performed instead.

In addition, participants (ie, students only) filled out a posttest survey (disclosed in Multimedia Appendix 2) in which their propensity toward the individual factors proposed by the social comparison model of competition [8] was assessed. Specifically, students' perception of closeness to their teachers, students' perception of closeness to their peers, students' perception of similarity to their teachers, students' perception of similarity to their peers, students' perception of relevance of the prescribed activities, and students' personality were assessed. To assess their personality, respondents completed a personality test in accordance with the Big Five personality traits [32]. The posttest survey was completed by 112 students.

Statistical Analysis

The first set of statistical analyses focused on the evaluation of dropouts. A respondent is labeled as a (provisional) dropout if the respondent has not visited the app in a given wave and is assumed to have lost interest (ie, dropped out) in the wave before. Several multiple regression models were fit to determine whether the number of dropouts changed over time and were different for each treatment. Subsequent analyses were performed on a subset of respondents who have participated in the study since the start of the first wave.

The second set of analyses focused on the evaluation of the engagement levels of both students and teachers. To evaluate treatment differences, further analyses were performed on respondents who actually had a chance to get exposed to the treatment. Therefore, from the entire data set, a subset was derived by preserving the combination of a particular respondent and wave only if the respondent had ever checked the app in that wave. Subsequently, several hierarchical linear models were estimated for the two outcome variables (ie, the number of days a participant had visited the app and the number of unique activities a respondent had performed) using time (ie, wave number), holiday, and treatment as predictors. We tested whether significant second-order interaction effects existed among these variables. In all models, we allowed random intercepts for both individuals and the classes they were in. The final model was selected based on Akaike information criterion [35]. Subsequently, in the final models, posthoc tests using Tukey adjustments were performed on the treatment variable. The same procedure was repeated to separately fit a model for students and teachers.

Finally, the third set of analyses focused on evaluating the impact of individual factors on engagement levels. Data on individual factors were derived from a posttest survey that was filled out by students only and not by teachers. Survey observations were linked via email addresses to GameBus user accounts to match individual factors with engagement levels. Although 112 students completed the survey, 10 responses could not be traced back to actual users of the mHealth platform. Furthermore, to evaluate the impact of individual factors, analyses were again performed on respondents who had a chance to get exposed to the treatment. Therefore, from the entire data set, a subset was derived, preserving the particular respondents that had checked the app over the course of the entire campaign at least twice, leaving 67 respondents in the data set for further analyses. Note that, in contrast to the second set of analyses, data are now aggregated over the course of the entire campaign and not per wave. Subsequently, several multiple regression models were fitted for the two outcome variables using the 10 individual factors as predictors. On the basis of Akaike information criterion, a backward selection procedure was used to select the final model [35].

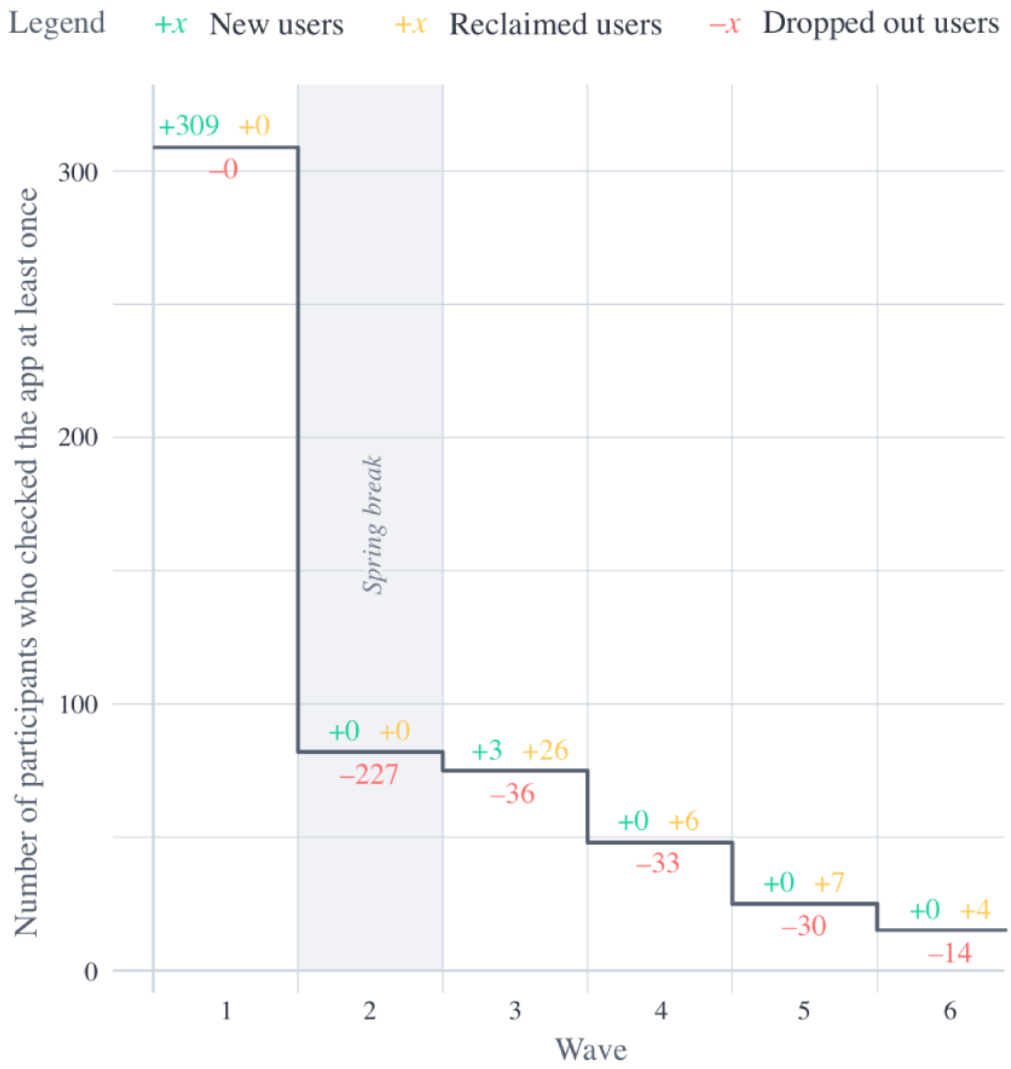
Results

User Statistics

In total, 313 unique participants, including 290 students (92.7%) and 23 teachers (7.3%), participated in the study. Educational level A included 61 students and 6 teachers, educational level B included 110 students and 9 teachers, and educational level C included 119 students and 8 teachers. Figure 2 displays the degradation of the number of participants who checked the mobile app during a given wave. The number of participants who joined the campaign for the first time in a given wave is displayed in green. The number of participants that dropped out in a given wave is shown in red. The number of participants who checked the mobile app in a given wave, although they dropped out in an earlier wave (ie, reclaimed users), is displayed in yellow. It was found that students tended to drop out, especially at the beginning of the campaign (ie, the wave number was significant at P=.003) and during holidays (P=.04). No significant differences in dropout rates within treatments were detected. Therefore, it is assumed that dropouts are spread equally over treatments.

XSL•FO
RenderX

Figure 2. Number of participants who checked the app (at least once) per wave.



Evaluation Outcomes

A total of 3 respondents only joined in the third wave and, therefore, were excluded from further statistical analysis, leaving a total of 99% (288/290) of students and 95.5% (22/23) of teachers in the data set.

Impact of the Situational Factors on Engagement Levels

Impact on the Average Number of Days Visiting the App

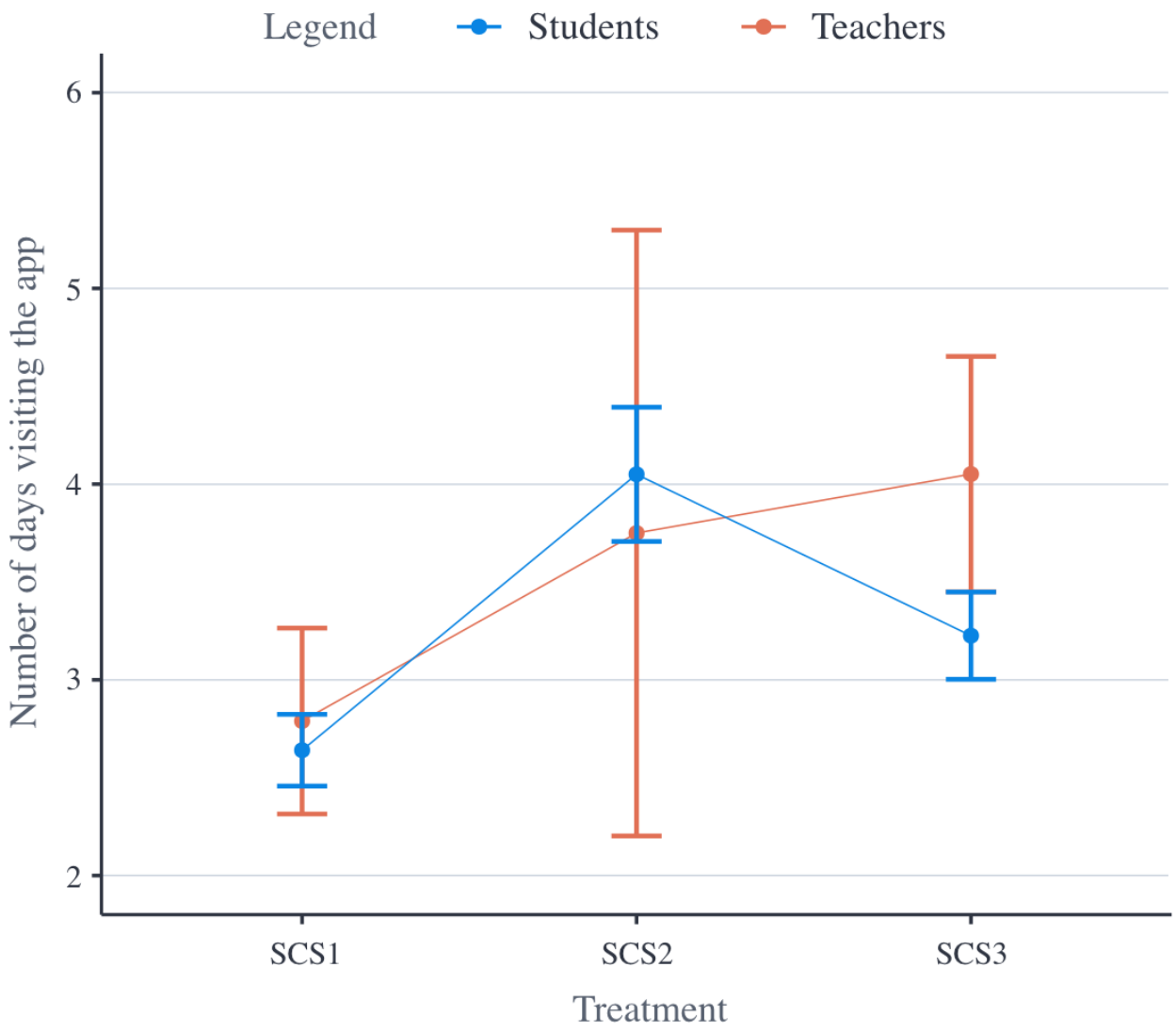
Figure 3 displays the number of days participants were visiting the app on average, per treatment. Note that the sloped horizontal lines are a visual aid to highlight the differences between the treatment group averages.

From the second set of statistical analyses, it was found that students' engagement dropped over time (ie, −0.740 days visiting the app per wave; $P<.001$) and also declined during holidays (ie, −0.595 days visiting the app; $P<.001$). In addition, it was found that students in SCS2 were significantly ($P<.001$) visiting the app more often (ie, +0.469 days) than students in SCS1. Further, students in SCS3 were visiting the app more often (ie, +0.215 days) than students in SCS1; however, this difference was not significant. Finally, students in SCS2 were visiting the app more often (ie, +0.255 days) than students in SCS3; however, this difference was not significant. No significant interaction effects were detected, and all treatments were equally impacted by holidays and time.

From the second set of statistical analyses, it was also found that teachers' engagement also decreased over time (ie, −0.763 days visiting the app per wave; $P<.001$). No additional significant (interaction) effects were observed. Teachers in SCS2 seemed to be visiting the app more often (ie, +1.250 days) than teachers in SCS1, teachers in SCS1 seemed to be visiting the app more often than teachers in SCS3 (ie, +0.173 days), and teachers in SCS3 also seemed to be visiting the app less often (ie, −1.423) than teachers in SCS2; however, none of these differences were reported to be significant.

XSL•FO
RenderX

Figure 3. Mean plot of the number of days participants have been visiting the app per treatment. SCS1: first social comparative setting; SCS2: second social comparative setting; SCS3: third social comparative setting.



Impact on the Average Number of Unique Activities

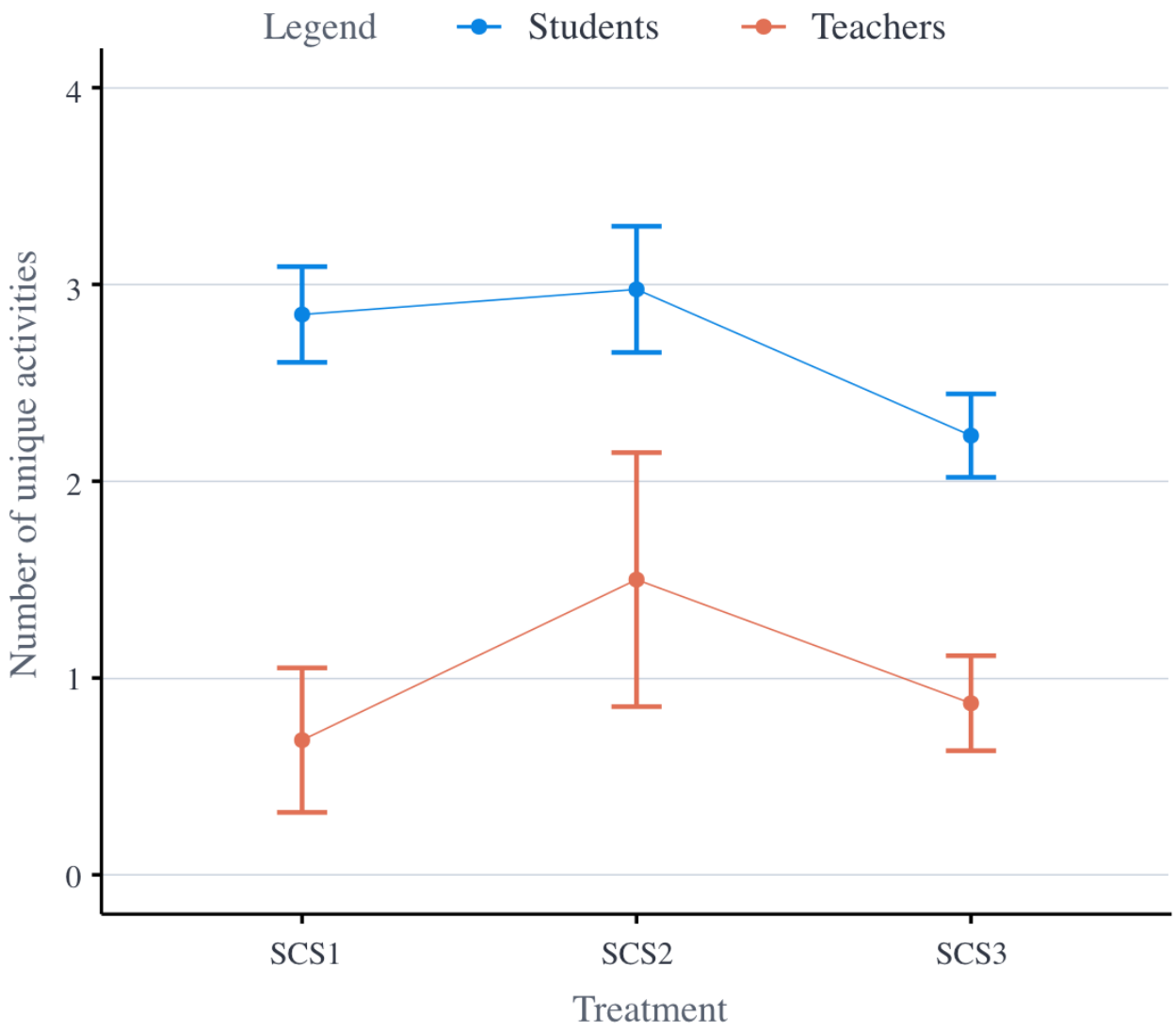
Figure 4 displays the average number of unique activities that participants performed per treatment. Multimedia Appendix 3 displays an overview of the number of times students and teachers performed each prescribed activity.

From the second set of statistical analyses, it was also found that the number of activities students performed decreased over time (ie, −1.249 activities per wave; $P<.001$) and during holidays (ie, −2.611 activities; $P<.001$). In addition, it was found that students in SCS2 performed, on average, more unique activities than students in SCS1 (ie, +0.457 activities); however, this difference was not significant. Students in SCS3, on the other hand, performed fewer unique activities, on average, than students in SCS1 (ie, −0.611 activities); however, this difference was not significant. Students in SCS2 performed significantly more activities, on average, than students in SCS3 (ie, +1.068 activities; $P=.004$). No significant interaction effects were detected; all treatments were equally impacted by holidays and time (ie, wave number).

From the second set of statistical analyses, it was found that the number of activities teachers performed also decreased over time (ie, −0.067 activities per wave; $P<.001$). Teachers, on average, performed fewer unique activities during holidays (ie, −0.019 activities); however, this difference was not significant. In addition, teachers in SCS2 performed more unique activities than teachers in SCS1 (ie, +1.227 activities; $P=.04$); teachers in SCS3 seemed to have performed slightly more unique activities than teachers in SCS1 (ie, +0.035 activities), although this difference was not significant; and teachers in SCS2, on average, performed more unique activities than teachers in SCS3 (ie, +1.192 activities; $P=.09$), although this difference was only close to significance.

XSL•FO
RenderX

Figure 4. Mean plot of the number of unique activities participants have performed per treatment. SCS1: first social comparative setting; SCS2: second social comparative setting; SCS3: third social comparative setting.



Impact of the Individual Factors on Student Engagement Levels

Impact on the Average Number of Days Visiting the App

To analyze the impact of individual factors on the average number of days participants checked the app, the model selection procedure selected a final model with four predictors. From this final model, it was found that students' perception of closeness to their teachers had a positive impact on the average number of days they were visiting the app (ie, +1.162 days; *P*=.03). In addition, it was found that the personality traits conscientiousness (+1.444 days; *P*=.02) and openness to experience (+1.398; *P*=.09) had a positive impact on the number of days students checked the app, although the impact of the latter personality trait was not significant. Finally, it was found that the personality trait extraversion had a negative impact on the number of days students were visiting the app (ie, −0.743 days; *P*=.03).

Impact on the Average Number of Unique Activities

To analyze the impact of individual factors on the average number of unique activities performed by participants, the model selection procedure selected a final model with three predictors. In particular, it was found that the students' perception of closeness to their teachers had a positive impact on the number of unique activities they performed (ie, +0.095 activities; *P*=.002). On the other hand, it was also found that the students' perception of closeness to their peers had a negative impact on the number of unique activities they performed (ie, −0.103 activities; *P*=.002). Finally, it was found that the students' perception of the relevance of the prescribed activities had a positive impact on the number of unique activities they performed (ie, +0.047 activities; *P*=.05).

Discussion

Principal Findings

The aim of this study is to evaluate the implementation of social comparison (ie, either fostering intra- or intergroup comparisons)

XSL•FO
RenderX

that is most effective in engaging preadolescent students in a health promotion campaign. Overall, our results indicated that students seem to monitor the intergroup competitions (ie, SCS2 and SCS3) more closely than the intragroup competition (ie, SCS1), as they, on average, checked the mHealth app more often when engaged in team-based comparisons. This result supports our hypothesis that an intergroup competition, with both its collaborative and competitive aspects, can better sustain engagement of students with an mHealth intervention than an intragroup competition, which involves only competitive aspects. In an intergroup competition, active players are more likely to discuss the position of their class on the leaderboard and encourage other class members to engage in the competition, as their own success (ie, winning the competition) depends on the performance of their class members. On the other hand, in an intragroup competition, it would have actually been beneficial for a student to be the only active player, as only the absolute winner would receive a small gift. Therefore, active players in the intragroup competition had no incentive to encourage other players to engage in the competition. Finally, the number of competitors was lower in the intergroup competitions than in the intragroup competition, which may have intensified the competition (ie, according to the N effect [21,22]).

In addition, it was found that students did, on average, complete (significantly) more unique activities in SCS2 (compared with SCS1 and SCS3). However, students in SCS3 completed fewest unique activities on average, whereas we expected that SCS3 would trigger, on average, the highest adoption of healthy routines, as we had introduced an additional comparative element in this setting (ie, by explicitly highlighting the performance of teachers).

In summary, we found that students adopted the most healthy routines in an intergroup competition and the fewest healthy routines in an intergroup competition. As a result, it is difficult to explain the difference between SCS2 and SCS3 based on the number of competitors or incentive structure because these were similar in both cases. However, this unexpected result may be explained by other factors. On the one hand, this result may be explained by the locked variable *educational level*. As we had to select a randomized block approach (ie, based on students' educational level) to distribute our treatments and given that the majority of our data were collected in the first wave (eg, due to increasing dropout over time), the crossover study design may not have prevented that the impact of a certain treatment is bound to a specific educational level.

On the other hand, the difference between the average number of unique activities that students performed in SCS2 and SCS3 may be explained by examining the teacher's performance in more detail; a plausible interpretation is that an intergroup competition, in which students cooperate with their mentoring teachers to beat other classes, requires actual involvement of the same mentoring teachers. The discrepancy between SCS2 and SCS3 may then be explained by the fact that teachers completed (significantly) more unique activities, on average, in SCS2 than in SCS3 and SCS1. The fluctuations in the number of activities teachers performed were not controlled for; however, in a social comparative setting where students are likely to draw inspiration from their teacher's actions [20], these coincidences can have an effect. By coincidence, teachers did not perform many activities in SCS1. Still, in this intragroup competition, students were probably sufficiently motivated by other active students to perform the healthy routines. It so happened that teachers were also passive in setting SCS3; although in this intergroup setting, their behavior probably demotivated students (who depended on them to pull up the class average and inspire other passive students in their class). It so happened that, compared with those in both SCS1 and SCS3, the teachers in SCS2 were actually, on average, performing (significantly) more unique activities. As a result, these teachers could have inspired their students, which explains the higher number of unique activities students performed on average.

Furthermore, when evaluating the individual factors that have influenced students' engagement levels, it was found that students' perception of closeness to their teachers had a positive impact on the average number of days they were visiting the app and the average number of unique activities they completed. It is likely that students who *feel closer to* their teachers participate more actively because their teachers have especially invited them to participate. This result supports the claim that teachers potentially serve as positive role models for students [14].

On the other hand, it was also found that students' perception of closeness to their peers had a negative impact on the number of unique activities they performed. Furthermore, in accordance with the findings of Beach and Tesser [25], it was found that students' perception of the relevance of the prescribed activities had a positive impact on the average number of unique activities students performed. It was also found that the personality trait conscientiousness had a positive impact on the average number of days students checked the mHealth app. However, in contrast to the findings of Buunk and Gibbons [33], it was found that the personality trait openness to experience actually had a positive impact on the number of days students checked the app. This may be explained by the context in which our study was executed; the health promotion campaign was advertised as a rather alternative form of education and is, therefore, likely perceived by students as *something new*. Finally, it was found that the personality trait extraversion had a negative impact on the number of days students were visiting the app. The negative impact of extraversion on students' engagement may be explained by the observation that extraverts are more easily bored [36] and, therefore, quit the mHealth intervention earlier than introverts.

Finally, it must be noted that engagement levels with the intervention dropped faster over time than expected. The spring break seemed to have a dramatic impact on students' engagement levels. In addition, teachers seemed to have been unable to drag their students' attention back to the health promotion campaign after the holiday period, although their role was implied to be important in raising campaign awareness.

Limitations

This study did not actively control fluctuations in the engagement levels of teachers (eg, the number of activities they performed). The diverse behavior of teachers has likely

XSL•FO
RenderX

influenced the engagement levels of students to some extent. In addition, as the focus of this study was on students' engagement, teachers did not fill out the posttest survey, which means that no qualitative data were collected on how they perceived the different treatments. Finally, no data were recorded on the number of reviews of the app's leaderboard teachers had actually hosted during plenary class settings.

Furthermore, although students did fill out the posttest survey, the degree of social relationships between the students was unclear at the start of the experiment. As a result, we could not assess what preadolescents were befriended, what students were most popular, and what teachers were beloved. Potentially, this analysis could have helped to target the most influential subjects and drag their attention back to the health promotion campaign after the spring break. Presumably, the most influential subjects could have also triggered the others to continue active participation.

Another weakness of this study is that social comparison was not studied in complete isolation (eg, some external rewards were provided as well). We kept the additional incentives stable across the treatment groups. Still, it is interesting to evaluate social comparison without any other incentives (eg, without the small gifts that were distributed in this study) to obtain a better estimate of the true impact social comparison has on engagement levels with an mHealth app.

Similarly, this study evaluated the impact of our intervention on a particular target group (ie, preadolescents) within a specific context (ie, the school environment). It is likely that the results will translate to other audiences and contexts because social comparison and its derivatives are universal processes [6-8]; however, it remains unclear what its impact on health behavior would be in different settings.

Future Work

A follow-up study should control the engagement levels of teachers (eg, by controlling the number of activities they perform) to analyze the exact impact of either passive or active teachers in intragroup and intergroup competitions. Further research is also needed to evaluate whether teachers are sufficiently strong positive role models for preadolescents. It has been demonstrated that social media influencers can serve as alternative role models by, for example, enhancing the dissemination of public health messages [37]. Therefore, a follow-up study could potentially benefit from social media influencer involvement. Finally, future studies may have students create their own teams (eg, in intergroup competitions). It was observed that people interact with different social networks (eg, a network of people for physical interaction and a network of people for sharing web-based messages) [38]. It would be interesting to investigate what type of network (ie, social comparative setting) is most effective in promoting the adoption of health routines among preadolescents.

Finally, we encourage studies evaluating persuasive strategies other than social comparison in this target population, such that we can compare the impact of individual behavior change techniques on preadolescents. We suggest that scholars should also conduct these studies within a relatively safe environment for preadolescents, such as their high schools. Although mHealth tools are deployed to promote something good in its users (ie, a person's health), the persuasive nature of behavior change techniques may potentially threaten an individual's freedom of conduct. Ethical guidelines for developing moral mHealth tools for preadolescents are still in their infancy. Dedicated research within the ethics community is trying to answer questions on the moral aspects of the development of mHealth tools [39], and guidelines for developing moral artificial intelligence interventions are emerging [40], which may also apply to specific mHealth interventions. We welcome additional ethical guidelines for the development of mHealth tools and the execution of empirical studies to evaluate these tools.

Conclusions

When using social comparison to engage preadolescents in a health promotion campaign using an mHealth tool, an intergroup competition—rather than an intragroup competition—can increase preadolescents' passive engagement with mHealth apps. However, an intergroup competition, as compared with an intragroup competition, does not necessarily result in preadolescents performing more unique activities on average. The active involvement of a role model (eg, a teacher) can influence the average number of unique activities preadolescents perform in an intergroup setting. For example, if the role model is active, preadolescents seems more likely to actively participate as well, because preadolescents are likely to draw inspiration from the actions of their role models. Moreover, preadolescents' perception of closeness to their role model seems to amplify this effect.

From this study, it is concluded that HPS interventions can use social dynamics to engage preadolescent students in healthier routines. However, additional behavior change strategies seem necessary to sustain students' engagement over time. In this process, an especially important role seems reserved for the teachers who serve as role models for their students and can potentially inspire them if they are actively involved in the HPS intervention themselves as well.

Acknowledgments

This work was part of the research program *Gamification for Overweight Prevention and Active Lifestyle* (443001101), which was partly financed by the Netherlands Organization for Health Research and Development.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Posttest survey.
[PDF File (Adobe PDF File), 397 KB - jmir_v23i7e21202_app1.pdf]

Multimedia Appendix 2
An overview of prescribed activities per wave.
[PDF File (Adobe PDF File), 69 KB - jmir_v23i7e21202_app2.pdf]

Multimedia Appendix 3
An overview of the number of times different prescribed activities were performed.
[PDF File (Adobe PDF File), 340 KB - jmir_v23i7e21202_app3.pdf]

References

1. Walsh JC, Groarke JM. Integrating behavioral science with mobile (mHealth) technology to optimize health behavior change interventions. Eur Psychol 2019 Jan;24(1):38-48. [doi: 10.1027/1016-9040/a000351]
2. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
3. Meeting Report: Nurturing Human Capital Along the Life Course: Investing in Early Child Development. World Health Organization. 2013. URL: https://apps.who.int/iris/handle/10665/87084 [accessed 2020-10-28]
4. Consideration of the Evidence on Childhood Obesity for the Commission on Ending Childhood Obesity: Report of the Ad Hoc Working Group on Science and Evidence for Ending Childhood Obesity. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/206549 [accessed 2020-10-28]
5. Hoorens V, Damme C. What do people infer from social comparisons? Bridges between social comparison and person perception. Soc Personal Psychol Compass 2012;6(8):607-618. [doi: 10.1111/j.1751-9004.2012.00451.x]
6. Gibbons FX, Buunk BP. Individual differences in social comparison: development of a scale of social comparison orientation. J Pers Soc Psychol 1999 Jan;76(1):129-142. [doi: 10.1037//0022-3514.76.1.129] [Medline: 9972558]
7. Festinger L. A theory of social comparison processes. Hum Relat 2016 Apr 22;7(2):117-140. [doi: 10.1177/001872675400700202]
8. Garcia SM, Tor A, Schiff TM. The psychology of competition: a social comparison perspective. Perspect Psychol Sci 2013 Nov;8(6):634-650. [doi: 10.1177/1745691613504114] [Medline: 26173228]
9. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
10. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining 'Gamification'. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek'11; September 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
11. Huotari K, Hamari J. Defining Gamification: a Service Marketing Perspective. In: Proceeding of the 16th International Academic MindTrek Conference. 2012 Presented at: MindTrek'12; October 3-5, 2012; Tampere, Finland p. 17-22. [doi: 10.1145/2393132.2393137]
12. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In: Proceedings of the 47th IEEE Hawaii international conference on system sciences. 2014 Presented at: HICSS'14; January 6-9, 2014; Waikoloa, HI, USA p. 3025-3034. [doi: 10.1109/hicss.2014.377]
13. Romero M. Learner engagement in the use of individual and collaborative serious games. Cutting-Edge Technol High Educ 2012;6:15-34 [FREE Full text] [doi: 10.1108/S2044-9968(2012)000006C004]
14. Orji R. Why Are Persuasive Strategies Effective? Exploring the Strengths and Weaknesses of Socially-oriented Persuasive Strategies. In: Proceedings of the 12th International Conference on Persuasive Technology. 2017 Presented at: PERSUASIVE'17; April 4-6, 2017; Amsterdam, The Netherlands p. 253-266. [doi: 10.1007/978-3-319-55134-0_20]
15. Bryant BK. The effects of the interpersonal context of evaluation on self- and other-enhancement behavior. Child Dev 1977 Sep;48(3):885. [doi: 10.2307/1128337]
16. Crockenberg SB, Bryant BK, Wilce LS. The effects of cooperatively and competitively structured learning enironments on inter- and intrapersonal behavior. Child Dev 1976 Jun;47(2):386-396. [Medline: 1269310]
17. Marker AM, Staiano AE. Better together: outcomes of cooperation versus competition in social exergaming. Games Health J 2015 Feb;4(1):25-30 [FREE Full text] [doi: 10.1089/g4h.2014.0066] [Medline: 26181677]
18. Wills TA. Downward comparison principles in social psychology. Psychol Bull 1981;90(2):245-271. [doi: 10.1037/0033-2909.90.2.245]

XSL•FO
RenderX

19. Smith RH. Assimilative and contrastive emotional reactions to upward and downward social comparisons. In: Suls J, Wheeler L, editors. Handbook of Social Comparison: Theory and Research. Boston, MA, USA: Springer; 2000. [doi: 10.1007/978-1-4615-4237-7_10]
20. Gibbons FX, Buunk BP. Superstars and me: predicting the impact of role models on the self. J Pers Soc Psychol 1997;73:91-103 [FREE Full text] [doi: 10.1037/0022- 3514.73.1.91]
21. Garcia SM, Tor A. The N-effect: more competitors, less competition. Psychol Sci 2009 Jul;20(7):871-877. [doi: 10.1111/j.1467-9280.2009.02385.x] [Medline: 19538434]
22. Tor A, Garcia SM. The N-effect. Psychol Sci 2010 Apr 29;21(5):748-749. [doi: 10.1177/0956797610368813]
23. Lewis ZH, Swartz MC, Lyons EJ. What's the point?: a review of reward systems implemented in gamification interventions. Games Health J 2016 Apr;5(2):93-99. [doi: 10.1089/g4h.2015.0078] [Medline: 26812253]
24. Nuijten RC, van Gorp PM, Kaymak U, Simons M, Kemperman AD, van den Berg PE. Evaluation of the impact of extrinsic rewards on user engagement in a health promotion context. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3600-3604. [doi: 10.1109/EMBC.2019.8856296] [Medline: 31946656]
25. Beach SR, Tesser A. Self-evaluation maintenance and evolution. In: Suls J, Wheeler L, editors. Handbook of Social Comparison. Boston, MA: Springer; 2000. [doi: 10.1007/978-1-4615-4237-7_7]
26. Tesser A, Campbell J. Self-evaluation maintenance and the perception of friends and strangers. J Personality 1982 Sep;50(3):261-279. [doi: 10.1111/j.1467-6494.1982.tb00750.x]
27. Tesser A. Toward a self-evaluation maintenance model of social behavior. Adv Exp Soc Psychol 1988;21:181-227. [doi: 10.1016/S0065-2601(08)60227-0]
28. Goethals GR, Darley JM. Social comparison theory: self-evaluation and group life. In: Theories of Group Behavior. New York, USA: Springer; 1987:21-47. [doi: 10.1007/978-1-4612-4634-3_2]
29. Houston JM, Mcintire SA, Kinnie J, Terry C. A factorial analysis of scales measuring competitiveness. Educ Psychol Meas 2016 Jul 2;62(2):284-298. [doi: 10.1177/0013164402062002006]
30. Darnon C, Dompnier B, Poortvliet P. Achievement goals in educational contexts: a social psychology perspective. Soc Personal Psychol Compass 2012;6(10):760-771. [doi: 10.1111/j.1751-9004.2012.00457.x]
31. Poortvliet PM, Janssen O, Van Yperen NW, Vliert EV. The joint impact of achievement goals and performance feedback on information giving. Basic Appl Soc Psychol 2009 Aug 6;31(3):197-209. [doi: 10.1080/01973530903058276]
32. Goldberg LR. An alternative 'description of personality': the big-five factor structure. J Pers Soc Psychol 1990 Dec;59(6):1216-1229. [doi: 10.1037//0022-3514.59.6.1216] [Medline: 2283588]
33. Buunk AP, Gibbons FX. Social comparison: the end of a theory and the emergence of a field. Organ Behav Hum Decis Process 2007 Jan;102(1):3-21. [doi: 10.1016/j.obhdp.2006.09.007]
34. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mhealth engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]
35. Grove D, Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. Stat 1988;37(4/5):477. [doi: 10.2307/2348776]
36. Geen RG. Preferred stimulation levels in introverts and extroverts: effects on arousal and performance. J Pers Soc Psychol 1984 Jun;46(6):1303-1312. [doi: 10.1037/0022-3514.46.6.1303]
37. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
38. van Woudenberg TJ, Bevelander KE, Burk WJ, Smit CR, Buijs L, Buijzen M. Comparing the measurement of different social networks: peer nominations, online communication, and proximity data. Net Sci 2020 Jan 31;8(1):62-78. [doi: 10.1017/nws.2019.65]
39. Frank LE. What do we have to lose? Offloading through moral technologies: moral struggle and progress. Sci Eng Ethics 2020 Feb;26(1):369-385 [FREE Full text] [doi: 10.1007/s11948-019-00099-y] [Medline: 30900160]
40. Ethics Guidelines for Trustworthy AI. The European Digital Strategy. 2019. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [accessed 2020-10-05]

Abbreviations

HPS: Health Promoting School
mHealth: mobile health
SCS1: first social comparative setting
SCS2: second social comparative setting
SCS3: third social comparative setting
VMBO: Voorbereidend middelbaar beroepsonderwijs

XSL•FO
RenderX

Edited by R Kukafka; submitted 10.06.20; peer-reviewed by V Visch, S Edney; comments to author 11.07.20; revised version received 28.10.20; accepted 24.05.21; published 29.07.21.

Please cite as:
Nuijten RCY, Van Gorp P, Borghouts T, Le Blanc P, Van den Berg P, Kemperman A, Hadian E, Simons M
Preadolescent Students' Engagement With an mHealth Intervention Fostering Social Comparison for Health Behavior Change: Crossover Experimental Study
J Med Internet Res 2021;23(7):e21202
URL: https://www.jmir.org/2021/7/e21202
doi:10.2196/21202
PMID:34326041

©Raoul Ceasar Yannic Nuijten, Pieter Van Gorp, Tom Borghouts, Pascale Le Blanc, Pauline Van den Berg, Astrid Kemperman, Ehsan Hadian, Monique Simons. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Guided Relaxation–Based Virtual Reality for Acute Postoperative Pain and Anxiety in a Pediatric Population: Pilot Observational Study

Vanessa A Olbrecht[1], MD, MBA; Keith T O'Conor[2], MD; Sara E Williams[3], PhD; Chloe O Boehmer[2], BA, BS; Gilbert W Marchant[2], MD; Susan M Glynn[2], BS; Kristie J Geisler[2], BS; Lili Ding[4], PhD; Gang Yang[4], MSc; Christopher D King[3], PhD

[1]Center for Understanding Pediatric Pain, Department of Anesthesiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[2]Department of Anesthesiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[3]Center for Understanding Pediatric Pain, Division of Behavioral Medicine and Clinical Psychology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[4]Division of Biostatistics and Epidemiology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

Corresponding Author:
Vanessa A Olbrecht, MD, MBA
Center for Understanding Pediatric Pain
Department of Anesthesiology
Cincinnati Children's Hospital Medical Center
3333 Burnet Avenue
MLC 2001
Cincinnati, OH, 45229
United States
Phone: 1 513 636 4408
Fax: 1 513 636 7337
Email: vanessa.olbrecht@cchmc.org

Abstract

Background: Distraction-based therapies, such as virtual reality (VR), have been used to reduce pain during acutely painful procedures. However, distraction alone cannot produce prolonged pain reduction to manage sustained postoperative pain. Therefore, the integration of VR with other pain-reducing therapies, like guided relaxation, may enhance its clinical impact.

Objective: The goal of this pilot study was to assess the impact of a single guided relaxation–based VR (VR-GR) session on postoperative pain and anxiety reduction in children. We also explored the influence of pain catastrophizing and anxiety sensitivity on this association.

Methods: A total of 51 children and adolescents (7-21 years) with postoperative pain and followed by the Acute Pain Service at Cincinnati Children's Hospital were recruited over an 8-month period to undergo a single VR-GR session. Prior to VR, the patients completed 2 questionnaires: Pain Catastrophizing Scale for Children (PCS-C) and the Child Anxiety Sensitivity Index (CASI). The primary outcome was a change in pain intensity following the VR-GR session (immediately, 15 minutes, and 30 minutes). The secondary outcomes included changes in pain unpleasantness and anxiety.

Results: The VR-GR decreased pain intensity immediately (P<.001) and at 30 minutes (P=.04) after the VR session, but not at 15 minutes (P=.16) postsession. Reductions in pain unpleasantness were observed at all time intervals (P<.001 at all intervals). Anxiety was reduced immediately (P=.02) but not at 15 minutes (P=.08) or 30 minutes (P=.30) following VR-GR. Patients with higher CASI scores reported greater reductions in pain intensity (P=.04) and unpleasantness (P=.01) following VR-GR. Pain catastrophizing was not associated with changes in pain and anxiety.

Conclusions: A single, short VR-GR session showed transient reductions in pain intensity, pain unpleasantness, and anxiety in children and adolescents with acute postoperative pain. The results call for a future randomized controlled trial to assess the efficacy of VR-GR.

Trial Registration: ClinicalTrials.gov NCT04556747; https://clinicaltrials.gov/ct2/show/NCT04556747

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e26328) doi:10.2196/26328

KEYWORDS

virtual reality; guided relaxation–based virtual reality; pain; anxiety; acute pain; postoperative pain; pediatrics

Introduction

Ineffective postoperative pain management, defined as surgery-related pain resulting from tissue injury and muscle spasm, has long-term consequences, including increased morbidity, poorer physical functioning, longer recovery, and increased costs [1]. Despite the widespread use of multimodal analgesia, pediatric postoperative pain remains difficult to manage [2], increasing the risk of persistent postoperative pain [3]. Studies of pediatric postoperative patients identified a 20% incidence of persistent pain beyond that expected from surgery [4]. While 80% of the patients included in this study recovered within 1 month, 20% reported reduced quality of life [4].

Although multimodal analgesia protocols focus on regional analgesia and nonopioid medications, opioid usage remains ubiquitous in pain management. Children and adolescents are at particular risk of long-term opioid abuse—as few as 5 days of use increases this risk [5]. A study of opioid-naïve pediatric surgical patients found persistent opioid use in 4.8% of adolescents versus 0.1% in a matched, nonsurgical cohort [3]. Other studies found that more than 25% of patients transitioned to chronic opioid consumption [6].

There remains a critical need for novel, nonpharmacologic pain management strategies, like virtual reality (VR). VR provides an immersive, multisensory, 3D environment that modifies the experiences of reality and creates a “sense of presence.” This “sense of presence” makes VR an excellent distraction-based therapy [7]. The use of distraction–based VR (VR-D) has been used to reduce pain in acute procedural, postoperative, and labor pain management by redirecting attention (eg, distraction) [8-14]. These transient reductions are sufficient for short-term reductions in pain, but they are not sufficient to treat prolonged acute pain experiences [15,16], including postoperative pain.

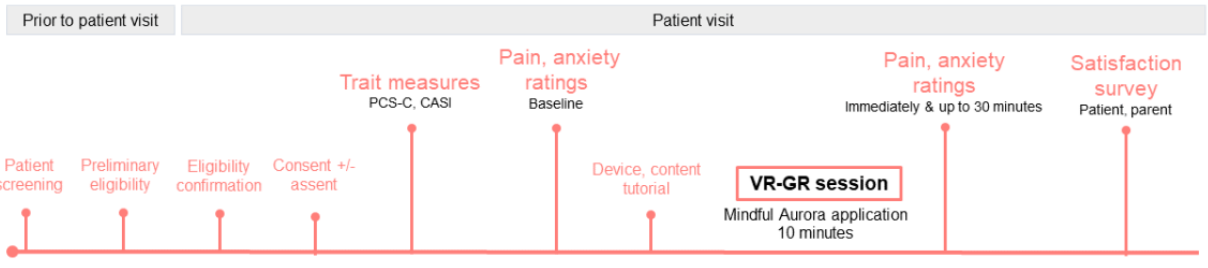
Alternative interventions utilizing traditionally delivered mind body–based therapies, like relaxation and slow breathing, reduce anxiety and pain in children undergoing surgery [17]. Combining these traditional mind-body therapies with VR, like guided relaxation–based VR (VR-GR), opens new possibilities for multimodal analgesia that may impact postoperative pain management and elevates VR therapy beyond simple distraction alone.

This study aimed to pilot the use of a single session of VR-GR in children after surgery to assess the association between VR-GR and pain intensity in addition to exploring the effect of pain catastrophizing and anxiety sensitivity on this association. We further assessed the association between VR-GR and reductions in pain unpleasantness and anxiety. We hypothesized that one VR-GR session would be associated with transient reductions in pain intensity, pain unpleasantness, and anxiety, with the greatest within-patient changes in children having high baseline pain catastrophizing and anxiety sensitivity.

Methods

A single-center, prospective, pilot study in a broad pediatric postoperative population experiencing moderate to severe pain was designed to preliminarily assess the correlation between a single VR-GR session and reductions in acute postoperative pain and anxiety, to explore the association between these within-patient changes and pain catastrophizing and anxiety sensitivity, and to show the feasibility of using this technology in a postoperative setting (Figure 1). Additionally, the study results will assist in developing a randomized controlled trial comparing the efficacy of VR-GR to active control, and will help in providing the preliminary effect data to assist with sample size determination.

Figure 1. Study diagram. CASI: Child Anxiety Sensitivity Index; PCS-C: Pain Catastrophizing Scale for Children; VR-GR: guided relaxation–based virtual reality.



Patients

The study recruited 51 children and adolescents who underwent surgery and were followed by the Acute Pain Service at Cincinnati Children’s Hospital Medical Center between July 2019 and March 2020. The study was approved by the Institutional Review Board (IRB #2018-2892) and conducted per the rules and regulations for ethical research. It was registered with ClinicalTrials.gov on September 21, 2020 (NCT04556747). A written consent (and assent for patients >11 years) was obtained from all participants and the study adhered to the CONSORT (Consolidated Standards of Reporting Trials) guidelines. The patients were recruited and enrolled in the study after surgery following hospital admission, and they were not compensated for participation.

XSL•FO
RenderX

Patients were suitable for recruitment if the Acute Pain Service was managing their pain. The patients managed by the Acute Pain Service experience high postoperative pain unmanageable solely by the surgery team. These patients necessitate multimodal analgesia and management with patient-controlled analgesia (PCA) and regional or neuraxial analgesia.

Inclusion Criteria

The eligibility criteria for inclusion were: 7-21 years old; able to read, speak, and write English; and undergone surgery resulting in significant postoperative pain necessitating care by the Acute Pain Service.

Exclusion Criteria

Patients were excluded if they did not fall within the age criteria; unable to read, speak, or write English; had a history of developmental delay, neurological conditions (seizure disorder, vertigo, dizziness, or significant motion sickness/nausea/vomiting), or uncontrolled psychiatric conditions; or had head or neck surgery that precluded the use of VR.

Patient Information

Prior to the VR-GR session, we collected the patient's age, sex, race, surgery type, and American Society of Anesthesiologists (ASA) status. The ASA status is a 6-category classification system that assesses a patient's general health before surgery; the sickness of a patient increases with the category number. Class I and II are healthy or have mild/moderate systemic disease; class III and IV patients have severe disease that may limit activity or threaten life.

Measures

The primary outcome was the change in pain intensity after VR-GR. Secondary outcome measures included changes in pain unpleasantness and anxiety.

Pain Intensity, Pain Unpleasantness, and Anxiety

Pain intensity, pain unpleasantness, and anxiety were measured before and after the VR-GR sessions using the numerical rating scale (NRS) [18]. A script was used to explain the difference between pain intensity and pain unpleasantness. Pain was described as analogous to listening to music on the radio: pain intensity as the music volume and pain unpleasantness as how much the patient disliked the music [19]. The patients rated symptom severity from 0 to 10 (0=nonexistent, 10=most severe) for each measure.

Pain Catastrophizing

The Pain Catastrophizing Scale for Children (PCS-C), a validated, self-reported questionnaire assessing pain catastrophizing tendencies [20,21], was completed before the VR-GR session. The survey was completed on an iPad by the patient directly into a REDCap database. All but one patient, who required that the survey be read to him or her, completed the survey independently. The patients used a 5-point Likert scale to rate 13 items assessing rumination, magnification, and helplessness related to thoughts about pain. The summary scores were interpreted as low=0-14, moderate=15-25, and high >26 [21]. Internal reliability in the sample was high (Cronbach α=.92).

Anxiety Sensitivity

The Child Anxiety Sensitivity Index (CASI) questionnaire was completed before the VR-GR as described above. The CASI has been used in VR studies in adolescents aged 10-21 years [12], validated in children, and is an 18-item survey that measures how patients perceive anxiety symptoms [22]. The total scores range from 18-54 [22], and the sample's internal reliability was good (Cronbach α=.86). Both surveys took less than 5 minutes to complete.

Patient Experience

Patient experience was measured with the study team–generated questionnaire. The participants ranked how much they agreed with statements on a 4-point Likert scale. The parent(s) of the participants were asked to fill out a similar survey to understand their perspective of the VR experience. The surveys, included as Multimedia Appendices 1 and 2, were completed independently on an iPad and entered directly into REDCap; 1 patient required that the survey be read aloud for completion. The surveys took less than 5 minutes to complete.

VR Device and Content

All patients used the Starlight Xperience VR device, a commercially available headset supplied by the Starlight Children's Foundation (Figure 2). It is a customized version of the Lenovo Mirage Solo with Daydream VR headset. The integrated headphones deliver audio content, and the patients interact and navigate within the VR environment using head movements and a handheld controller. The patients used the "Mindful Aurora" guided relaxation–based application to learn slow breathing and relaxation, an application developed by the Stanford CHARIOT (Childhood Anxiety Reduction through Innovation and Technology) program (Figure 3). The users were "transported" to an alpine meadow in the virtual world, where a 10-minute relaxation narrative teaches focused, slow, and paced breathing.

XSL•FO
RenderX

Figure 2. A child wearing a Starlight Xperience headset.



Figure 3. A snapshot of the Mindful Aurora application.



XSL•FO
RenderX

Procedures

The patient visits occurred on postoperative day (POD) 1 or 2. After determining eligibility and obtaining consent, the patients completed the PCS-C and CASI questionnaires and rated their baseline pain intensity, pain unpleasantness, and anxiety levels on the NRS.

Following these assessments, the patients were oriented to the VR headset and given a tutorial on the device application. They were instructed to remove the headset for discomfort, nausea, or dizziness. After completing the 10-minute session (based on the standard duration of relaxation sessions), the VR device was removed. Pain intensity, pain unpleasantness, and anxiety were recorded immediately, at 15 minutes, and at 30 minutes after the experience. The patients and parents then completed the experience questionnaires.

Statistical Analysis

All statistical analyses were performed using SAS 9.4 (SAS Institute). A *P* value of .05 was the cutoff value for statistical significance. Additionally, statistical significance with the Bonferroni adjustment for multiple comparisons for the primary outcome (change from the baseline pain intensity at three time-points after VR-GR) was assessed. The first-order autoregressive, AR(1), covariance structure was used in all the mixed-effects models. The missing data were examined, and all available data were used in statistical analyses.

Descriptive Analysis

Descriptive statistics were calculated for all baseline variables and change from the baseline for outcome variables. Mean (SD) and median (IQR) were used for the continuous variables, while frequency and percentage were used for the categorical variables.

Changes in Pain and Anxiety Following VR-GR

Pain intensity, unpleasantness, and anxiety after the VR-GR session were compared to the baseline within each patient using paired tests (*t* test or signed-rank, as appropriate) at individual time points. The changes from baseline were analyzed with mixed-effects models where time intervals (immediately, 15 minutes, or 30 minutes after VR-GR), pain catastrophizing, and anxiety sensitivity were used as the categorical fixed effects.

Associations in Baseline Outcomes

The Pearson or Spearman correlation coefficients were derived to test the association between the traits (pain catastrophizing and anxiety sensitivity) and outcomes (pain intensity, pain unpleasantness, and anxiety).

Impact of Psychological Factors on Changes in Pain and Anxiety

Mixed-effects models were used to examine the association of PCS-C and CASI on the changes from the baseline of pain intensity, pain unpleasantness, and anxiety, where time intervals (immediately, 15 minutes, or 30 minutes after the VR-GR) were used as the categorical fixed effect.

Results

Participant Characteristics

We enrolled 51 patients over 8 months. All patients completed pain and anxiety assessments at baseline and immediately following VR-GR; 100% (n=51) completed the pain assessment at 15 minutes, and 98% (n=50) completed the anxiety assessment at 15 minutes; and 88% (n=45) completed the assessments at 30 minutes following VR-GR. The patients did not receive any analgesic medications during participation in the study. Missing data resulted from limitations in the clinical environment, including patients undergoing imaging studies, receiving care from the care team, or falling asleep.

The patients were primarily adolescent, male, and Caucasian (Table 1). Of the 51 patients recruited, 19 (37.3%) underwent abdominal surgery, 21 (41.2%) underwent Nuss repair of pectus excavatum or chest surgery, and 11 (21.6%) underwent orthopedic procedures (such as posterior spinal fusion or major hip surgery). Half of the recruited patients were ASA status I/II or III/IV. The patients reported moderate pain intensity, unpleasantness, and mild anxiety levels before the VR-GR (Table 1).

Patients had moderate pain catastrophizing and average anxiety sensitivity [21]. Furthermore, higher PCS-C scores were associated with higher baseline anxiety (Spearman ρ=0.41, P<.001) (Table 2).

Pain catastrophizing and anxiety sensitivity were not associated with changes in pain intensity or pain unpleasantness.

XSL•FO
RenderX

Table 1. Demographic, survey, and medical data from study participants (N=51).

Variable	Value
Age (years)	14.6 (3.2)
Sex, n (%)	
Male	32 (63)
Female	19 (37)
Race, n (%)	
Caucasian	41 (80)
Non-Caucasian	10 (20)
Surgery type, n (%)	
Pectus/chest	21 (41)
Abdominal	19 (37)
Orthopedic	11 (22)
ASA[a] physical status, n (%)	
I/II (healthy, mild systemic disease)	24 (47)
III/IV (severe or life-threatening disease)	24 (47)
Baseline NRS[b] scores, n (%)	
Pain intensity (0-10)	5.11 (1.74)
Pain unpleasantness (0-10)	5.73 (2.30)
Anxiety (0-10)	2.05 (2.50)
Psychological factors, n (%)	
Catastrophizing (PCS-C[c])	21.6 (11.0)
Anxiety sensitivity (CASI[d])	31.2 (3.9)

[a]ASA: American Society of Anesthesiologists; ASA status was not collected on three (n=3) patients.

[b]NRS: numerical rating scale.

[c]PCS-C: Pain Catastrophizing Scale for Children.

[d]CASI: Child Anxiety Sensitivity Index.

Table 2. The Spearman correlation between baseline numerical rating scale (NRS) scores and surveys.

Baseline NRS	Pain catastrophizing (PCS-C[a])	*P* value	Anxiety sensitivity (CASI[b])	*P* value
Pain intensity	0.19	.17	–0.09	.55
Pain unpleasantness	0.17	.23	0.06	.67
Anxiety	0.41	.003	0.25	.07

[a]PCS-C: Pain Catastrophizing Scale for Children.

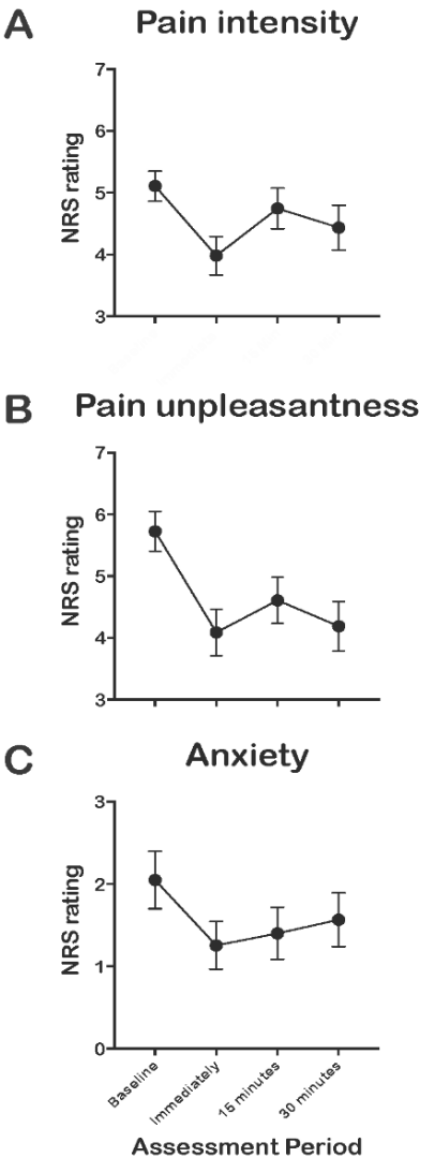
[b]CASI: Child Anxiety Sensitivity Index.

Primary Outcome: Pain Intensity

VR-GR was associated with a small reduction in pain intensity (Figure 4A, Table 3). The Wilcoxon signed-rank test showed that pain intensity decreased immediately (median –1.0, IQR –2.0 to 0, $P<.001$) following the VR-GR session and remained significant at 15 minutes (median 0, IQR –1.0 to 0.50, $P=.03$), and at 30 minutes (median 0, IQR –1.5 to 0, $P=.02$) (Table 3).

XSL•FO
RenderX

Figure 4. Changes in pain intensity, pain unpleasantness, and anxiety numerical rating scale (NRS) scores at baseline, immediately, 15 minutes, and 30 minutes. (A) The Wilcoxon signed-rank test showing the VR-GR correlation with pain intensity: immediately (median –1.0, IQR –2.0 to 0, *P*<.001); at 15 minutes (median 0, IQR –1.0 to 0.50, *P*=.03); at 30 minutes (median 0, IQR –1.5 to 0, *P*=.02). (B) The Wilcoxon signed-rank test showing the VR-GR correlation with pain unpleasantness: immediately (median –2.0, IQR –3.0 to 0, *P*<.001); at 15 minutes (median –1.0, IQR –2.0 to 0, *P*<.001); and at 30 minutes (median –1.0, IQR –3.0 to 0, *P*<.001). (C) The Wilcoxon signed-rank test showing an association between the VR-GR and anxiety: immediately (median 0, IQR –1.0 to 0, *P*<.001); at 15 minutes (median 0, IQR –1.0 to 0, *P*=.01); at 30 minutes (median 0, IQR –1.0 to 0, *P*=.17).



XSL•FO
RenderX

Table 3. Mean (SD) and median (IQR) changes in pain intensity, pain unpleasantness, and anxiety numerical rating scale (NRS) scores following the guided relaxation–based virtual reality (VR-GR). The Wilcoxon signed-rank test was used to compare the changes.

NRS outcome following the VR-GR	N	Mean (SD)	Median (IQR)	*P* value
Pain intensity				
Immediately	51	–1.1 (1.1)	–1 (–2.0 to 0)	<.001
15 minutes	51	–0.4 (1.5)	0 (–1.0 to 0.5)	.03
30 minutes	45	–0.5 (1.8)	0 (–1.5 to 0)	.02
Pain unpleasantness				
Immediately	51	–1.6 (1.6)	–2 (–3.0 to 0)	<.001
15 minutes	51	–1.1 (1.9)	–1 (–2.0 to 0)	<.001
30 minutes	45	–1.4 (1.9)	–1 (–3.0 to 0)	<.001
Anxiety				
Immediately	51	–0.8 (1.4)	0 (–1.0 to 0)	<.001
15 minutes	51	–0.6 (1.7)	0 (–1.0 to 0)	.01
30 minutes	45	–0.4 (1.7)	0 (–1.0 to 0)	.17

Secondary Outcomes

Pain Unpleasantness

VR-GR was also associated with a change in pain unpleasantness (Figure 4B). The Wilcoxon signed-rank test (Table 3) revealed that pain unpleasantness decreased immediately following VR-GR (median –2.0, IQR –3.0 to 0, *P*<.001), and remained significantly decreased at 15 minutes (median –1.0, IQR –2.0 to 0, *P*<.001), and at 30 minutes (median –1.0, IQR –3.0 to 0, *P*<.001) following VR-GR.

Anxiety

We also found an association between VR-GR and anxiety (Figure 4C). The Wilcoxon signed-rank test (Table 3) revealed anxiety decreased from the baseline to immediately following VR-GR (median 0, IQR –1.0 to 0, *P*<.001); it remained lower at 15 minutes (median 0, IQR –1.0 to 0, *P*<.001) but not at 30 minutes.

Association Between Psychological Factors and Changes in Outcomes From the Baseline

We used a series of adjusted, mixed-effects models to observe the correlation between psychological factors and changes in the baseline outcomes. It was observed that anxiety sensitivity levels were associated with greater reductions in pain intensity (β=–.06, SE 0.03, *P*=.04) and pain unpleasantness (β=–.09, SE 0.04, *P*=.01). Pain catastrophizing was not associated with changes in pain and anxiety following VR-GR (Table 4).

Table 4. Association between psychological factors and changes in pain intensity, pain unpleasantness, and anxiety numerical rating scale (NRS) scores following the guided relaxation–based virtual reality (VR-GR).

Change from the baseline following the VR-GR	Beta (SE)	*P* value
Pain intensity		
Pain catastrophizing (PCS-C[a])	–.0001 (0.01)	.99
Anxiety sensitivity (CASI[b])	–.04 (0.02)	.06
Pain unpleasantness		
Pain catastrophizing (PCS-C)	.01 (0.02)	.50
Anxiety sensitivity (CASI)	–.05 (0.03)	.08
Anxiety		
Pain catastrophizing (PCS-C)	–.02 (0.02)	.25
Anxiety sensitivity (CASI)	–.06 (0.03)	.04

[a]PCS-C: Pain Catastrophizing Scale for Children.

[b]CASI: Child Anxiety Sensitivity Index.

When assessing the association between changes from the baseline and covariates, we found that the Caucasian patients had a smaller decrease in pain unpleasantness from the baseline than the non-Caucasian participants—difference in least squares mean (LSM=1.20, 95% CI 0.11-2.28, *P*=.03). Additionally, the older participants had a smaller decrease in pain intensity (β=.12, SE 0.05, *P*=.03) than the younger patients. No other variables (ASA, POD, or race) were associated with changes in outcomes.

XSL•FO
RenderX

Satisfaction With VR

The patients reported a very positive VR experience. Overall, 96% of children would recommend VR to friends and family, and most (n=45, 88%) believed they felt "calmer and less anxious after having used VR" and that VR "made it easier for (them) to tolerate (their) pain." Parents reported a similar positive overall experience when asked the same questions. The parents (n=44, 100%) who completed the questionnaire would recommend VR; 93% (n=41) believed that VR made their child calm, and 84% (n=37) believed that VR helped their child tolerate pain. No patient experienced any self-reported side effects during the VR session.

Discussion

Principal Findings

VR-GR may be beneficial in pediatric postoperative pain management. Our pilot study assessed the association between a single VR-GR session delivered after surgery and changes in pain intensity, pain unpleasantness, and anxiety in children and adolescents with acute pain following surgery. The most significant changes were observed for pain unpleasantness, with lesser reductions in pain intensity and anxiety. More considerable reductions in pain intensity and unpleasantness after VR-GR were associated with higher anxiety sensitivity. Pain catastrophizing did not appear to be associated with changes in these outcomes. The qualitative assessments suggested that the patients had positive experiences with VR-GR with no reported side effects.

Although pain management strategies increasingly incorporate multimodal analgesia, the percentage of patients experiencing severe postoperative pain did not change in the last 20 years [2]. Besides prolonged pain, children are also at risk of persistent opioid use after surgery; opioid exposure after surgery can confer nearly a 50-fold increase in opioid use [3]. It was reported that more than 25% of children with chronic pain transitioned into chronic opioid use after using opioids for surgery-associated pain [6].

Few studies have used VR for acute postoperative pain in adults [23], and none have used it to help manage pain in children after surgery. Most VR studies use distraction–based VR (VR-D) to reduce pain temporarily by redirecting attention [7,9,10,13]. Without VR, distraction alone provides little pain management benefit [24], with no significant lasting impact on pain [15]. The improved efficacy of VR-D to temporarily reduce pain versus distraction alone may be due to the immersion with VR [23]. Although VR-D is more effective than distraction alone, its use is limited. Incorporating VR with other pain-reducing strategies, like guided relaxation, may promote sustained pain relief beyond the temporary impact of distraction [16].

Mind-body therapies, like relaxation and guided imagery, decrease anxiety and help manage pain in children undergoing surgery [17]. Unfortunately, their integration into postoperative clinical care is fraught with challenges, including access to care, high cost, and provider availability. As a result, the use of these therapies is minimal even at a tertiary care institution like ours, besides their inability to engage children. Using VR to deliver these therapies can increase accessibility and enhance acceptability in children versus methods without VR, making this therapy more engaging and relevant. Importantly, VR and mind-body therapies have very little side effects, a significant advantage over traditional pharmacologic interventions.

This study was the first to integrate a mind-body therapy with VR to help pediatric postoperative pain. The study's goals were to preliminarily assess the use of VR-GR to decrease acute postoperative pain and anxiety and determine whether pain catastrophizing and anxiety sensitivity influence the VR-GR effect in a broad pediatric population. The "Mindful Aurora" application taught patients to slow down their breathing and relax. A single VR-GR session was associated with immediate, acute changes in pain intensity, pain unpleasantness, and anxiety. Reductions in pain unpleasantness were observed at all intervals (immediately, 15 minutes, and 30 minutes following VR-GR). Anxiety reduced immediately and at 15 minutes following VR-GR. An association was observed between higher CASI scores and greater reductions in pain intensity and unpleasantness following VR-GR; however, such an association was absent with pain catastrophizing.

Limitations

This study had several limitations. The study design did not include a control group, which restricted us from determining a causal relationship between the use of VR-GR and pain and anxiety reduction. It is possible that interaction with the study team could influence changes in pain and anxiety. Additionally, we did not collect data on analgesic use in the study population or standardize the timing of the postoperative visit. This study was designed to test the feasibility of the VR-GR technology, obtain pilot data in children hospitalized after surgery, and assess a possible association between VR-GR and pain and anxiety reduction. The results of this study support the need to conduct a randomized controlled efficacy trial comparing the use of VR-GR to active control, data collection on the timing and use of all analgesic medications, and standardizing the timing of the VR-GR sessions. The study used a single, short VR session in the postoperative period. Although we saw small changes in pain intensity, pain unpleasantness, and anxiety, these changes did not reach any clinical significance. Previous literature report that a reduction in pain intensity of 2 or more points on the NRS or a 30% reduction in pain is considered clinically significant [25]. We recognize that a single session is unlikely to produce a large and sustained effect, and we will pursue further clinical studies using repeated sessions to determine the impact on these outcomes. Finally, the potential bias when self-reporting pain scores could have produced lower self-reported pain and anxiety scores following the VR-GR session. This will be addressed in future studies by including a control group.

Conclusions

To summarize, our study demonstrates the successful use of VR-GR in children after surgery. A single VR-GR session was associated with transient decreases in pain and anxiety. Future research is needed to investigate the effect of VR-GR in reducing pain and anxiety in the postoperative setting compared to a control.

XSL•FO
RenderX

Acknowledgments

The authors thank Maria Ashton, MS, RPh, MBA, medical writer, Department of Anesthesia, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, for writing assistance, editing, and proofreading. The authors would also like to thank Luke Wilson and the Stanford CHARIOT program for the development and use of the Mindful Aurora application.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient experience questionnaire - parent.
[PDF File (Adobe PDF File), 42 KB - jmir_v23i7e26328_app1.pdf]

Multimedia Appendix 2
Patient experience questionnaire - child.
[PDF File (Adobe PDF File), 48 KB - jmir_v23i7e26328_app2.pdf]

References

1. Gan TJ. Poorly controlled postoperative pain: Prevalence, consequences, and prevention. J Pain Res 2017;10:2287-2298 [FREE Full text] [doi: 10.2147/JPR.S144066] [Medline: 29026331]
2. Kozlowski LJ, Kost-Byerly S, Colantuoni E, Thompson CB, Vasquenza KJ, Rothman SK, et al. Pain prevalence, intensity, assessment and management in a hospitalized pediatric population. Pain Manag Nurs 2014 Mar;15(1):22-35. [doi: 10.1016/j.pmn.2012.04.003] [Medline: 24602421]
3. Harbaugh CM, Lee JS, Hu HM, McCabe SE, Voepel-Lewis T, Englesbe MJ, et al. Persistent opioid use among pediatric patients after surgery. Pediatrics 2018 Jan;141(1):e20172439 [FREE Full text] [doi: 10.1542/peds.2017-2439] [Medline: 29203521]
4. Rabbitts JA, Zhou C, Groenewald CB, Durkin L, Palermo TM. Trajectories of postsurgical pain in children: Risk factors and impact of late pain recovery on long-term health outcomes after major surgery. Pain 2015 Nov;156(11):2383-2389 [FREE Full text] [doi: 10.1097/j.pain.0000000000000281] [Medline: 26381701]
5. Shah A, Hayes CJ, Martin BC. Characteristics of initial prescription episodes and likelihood of long-term opioid use - United States, 2006-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 17;66(10):265-269 [FREE Full text] [doi: 10.15585/mmwr.mm6610a1] [Medline: 28301454]
6. Callinan CE, Neuman MD, Lacy KE, Gabison C, Ashburn MA. The initiation of chronic opioids: A survey of chronic pain patients. J Pain 2017 Apr;18(4):360-365. [doi: 10.1016/j.jpain.2016.11.001] [Medline: 27919771]
7. Carrougher GJ, Hoffman HG, Nakamura D, Lezotte D, Soltani M, Leahy L, et al. The effect of virtual reality on pain and range of motion in adults with burn injuries. J Burn Care Res 2009;30(5):785-791 [FREE Full text] [doi: 10.1097/BCR.0b013e3181b485d3] [Medline: 19692911]
8. Morris L, Louw Q, Grimmer-Somers K. The effectiveness of virtual reality on reducing pain and anxiety in burn injury patients: A systematic review. Clin J Pain 2009;25(9):815-826. [doi: 10.1097/AJP.0b013e3181aaa909] [Medline: 19851164]
9. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: A systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
10. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: Current trends and future directions. Pain Manag 2011 Mar;1(2):147-157 [FREE Full text] [doi: 10.2217/pmt.10.15] [Medline: 21779307]
11. Garrett B, Taverner T, Masinde W, Gromala D, Shaw C, Negraeff M. A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. Clin J Pain 2014 Dec;30(12):1089-1098. [doi: 10.1097/AJP.0000000000000064] [Medline: 24535053]
12. Gold JI, Mahrer NE. Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
13. Indovina P, Barone D, Gallo L, Chirico A, De Pietro G, Giordano A. Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: A comprehensive literature review. Clin J Pain 2018 Sep;34(9):858-877. [doi: 10.1097/AJP.0000000000000599] [Medline: 29485536]
14. Frey DP, Bauer ME, Bell CL, Low LK, Hassett AL, Cassidy RB, et al. Virtual reality analgesia in labor: The VRAIL pilot study - a preliminary randomized controlled trial suggesting benefit of immersive virtual reality analgesia in unmedicated laboring women. Anesth Analg 2019 Jun;128(6):e93-e96. [doi: 10.1213/ANE.0000000000003649] [Medline: 31094789]

XSL•FO
RenderX

15. Van Ryckeghem DM, Van Damme S, Eccleston C, Crombez G. The efficacy of attentional distraction and sensory monitoring in chronic pain patients: A meta-analysis. Clin Psychol Rev 2018 Feb;59:16-29. [doi: 10.1016/j.cpr.2017.10.008] [Medline: 29126746]
16. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: Does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 01;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
17. Vagnoli L, Bettini A, Amore E, De Masi S, Messeri A. Relaxation-guided imagery reduces perioperative anxiety and pain in children: A randomized study. Eur J Pediatr 2019 Jun;178(6):913-921. [doi: 10.1007/s00431-019-03376-x] [Medline: 30944985]
18. Tsze DS, von Baeyer CL, Pahalyants V, Dayan PS. Validity and reliability of the verbal numerical rating scale for children aged 4 to 17 years with acute pain. Ann Emerg Med 2018 Jun;71(6):691-702.e3 [FREE Full text] [doi: 10.1016/j.annemergmed.2017.09.009] [Medline: 29107409]
19. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983 Sep;17(1):45-56. [doi: 10.1016/0304-3959(83)90126-4] [Medline: 6226917]
20. Pagé GM, Campbell F, Isaac L, Stinson J, Martin-Pichora AL, Katz J. Reliability and validity of the Child Pain Anxiety Symptoms Scale (CPASS) in a clinical sample of children and adolescents with acute postsurgical pain. Pain 2011 Sep;152(9):1958-1965. [doi: 10.1016/j.pain.2011.02.053] [Medline: 21489692]
21. Pielech M, Ryan M, Logan D, Kaczynski K, White MT, Simons LE. Pain catastrophizing in children with chronic pain and their parents: Proposed clinical reference points and reexamination of the Pain Catastrophizing Scale measure. Pain 2014 Nov;155(11):2360-2367 [FREE Full text] [doi: 10.1016/j.pain.2014.08.035] [Medline: 25180013]
22. Silverman WK, Goedhart AW, Barrett P, Turner C. The facets of anxiety sensitivity represented in the Childhood Anxiety Sensitivity Index: Confirmatory analyses of factor models from past studies. J Abnorm Psychol 2003 Aug;112(3):364-374. [doi: 10.1037/0021-843x.112.3.364] [Medline: 12943015]
23. Mosso JL, Rizzo S, Wiederhold B, Lara V, Flores J, Espiritusanto E, et al. Cybertherapy--new applications for discomfort reductions. Surgical care unit of heart, neonatology care unit, transplant kidney care unit, delivery room-cesarean surgery and ambulatory surgery, 27 case reports. Stud Health Technol Inform 2007;125:334-336. [Medline: 17377297]
24. Eccleston C. Chronic pain as embodied defence: Implications for current and future psychological treatments. Pain 2018 Sep;159 Suppl 1:S17-S23. [doi: 10.1097/j.pain.0000000000001286] [Medline: 30113943]
25. Rowbotham M. What is a "clinically meaningful" reduction in pain? Pain 2001 Nov;94(2):131-132. [doi: 10.1016/S0304-3959(01)00371-2] [Medline: 11690725]

Abbreviations

ASA: American Society of Anesthesiologists
CASI: Child Anxiety Sensitivity Index
CONSORT: Consolidated Standards of Reporting Trials
CHARIOT: Childhood Anxiety Reduction through Innovation and Technology
LSM: least squares mean
NRS: numerical rating scale
PCA: patient-controlled analgesia
PCS-C: Pain Catastrophizing Scale for Children
POD: postoperative day
VR: virtual reality
VR-D: distraction–based virtual reality
VR-GR: guided relaxation–based virtual reality

Edited by R Kukafka; submitted 07.12.20; peer-reviewed by T Caruso, V Rocío, M Linnett; comments to author 22.01.21; revised version received 02.02.21; accepted 04.05.21; published 12.07.21.

Please cite as:
Olbrecht VA, O'Conor KT, Williams SE, Boehmer CO, Marchant GW, Glynn SM, Geisler KJ, Ding L, Yang G, King CD
Guided Relaxation–Based Virtual Reality for Acute Postoperative Pain and Anxiety in a Pediatric Population: Pilot Observational Study
J Med Internet Res 2021;23(7):e26328
URL: https://www.jmir.org/2021/7/e26328
doi:10.2196/26328
PMID:34048358

XSL•FO
RenderX

©Vanessa A Olbrecht, Keith T O'Conor, Sara E Williams, Chloe O Boehmer, Gilbert W Marchant, Susan M Glynn, Kristie J Geisler, Lili Ding, Gang Yang, Christopher D King. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effects of Immersive Virtual Reality in Reducing Public Stigma of Mental Illness in the University Population of Hong Kong: Randomized Controlled Trial

Anna S Y Yuen[1], BSocSci; Winnie W S Mak[1], PhD

Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Corresponding Author:
Winnie W S Mak, PhD
Department of Psychology
The Chinese University of Hong Kong
Room 354, Sino Building
Hong Kong
Hong Kong
Phone: 852 3943 6577
Email: wwsmak@cuhk.edu.hk

Abstract

Background: Public stigma in mental health often brings various adverse effects on people with mental illness. Researchers have been developing different interventions in combating public stigma.

Objective: This study investigates the effects of immersive virtual reality (IVR) in reducing the public stigma of mental illness using a single-blinded randomized control trial.

Methods: A pre-post experimental design with a 1-week follow-up was conducted. Participants (N=206) were recruited through the mass mail system of The Chinese University of Hong Kong and randomized into 3 conditions: immersive animation, text, and control. In the immersive animation condition (n=72), participants experienced the simulation of daily life and the stigma experienced as an animated story protagonist with mixed anxiety and depressive disorder with IVR. In the text condition (n=65), participants experienced an identical story to the immersive animation condition with first-person audio narration using the same virtual reality headset. In the control condition (n=69), participants watched a video about planets with IVR. All participants received interventions with a researcher-assisted Oculus Go virtual reality headset. Participants' public stigma was measured through self-administered online questionnaires and compared across conditions and at different time points using repeated measures analysis of variance. Simple and sequential mediation analyses on the relationship of condition (immersive animation vs text) and follow-up public stigma with possible mediators, including sense of embodiment and story transportation, were conducted using PROCESS.

Results: Public stigma did not differ significantly across conditions at pre-experiment (P>.99). In the immersive animation and text conditions, public stigma was significantly reduced at postexperiment and at the 1-week follow-up compared to pre-experiment (all with P<.001). Public stigma in the control condition at postexperiment and follow-up remained unchanged compared with pre-experiment (P=.69). Immersive animation had significantly lower public stigma than the control at postexperiment (P=.003) and follow-up (P=.02). Text also had lower public stigma than the control at postexperiment (P=.007) and follow-up (P=.03). However, immersive animation did not significantly differ from text in public stigma at postexperiment and follow-up (both P>.99). In simple mediation models, both sense of embodiment (95% CI –0.22 to 0.46) and story transportation (95% CI –0.18 to 0.00) were not significant mediators. In the sequential mediation model, both sense of embodiment and story transportation were significant sequential mediators. Sense of embodiment was positively associated with story transportation (P<.001), while story transportation was negatively associated with public stigma (P<.001). The indirect effect of the sequential mediation model was significant (95% CI –0.38 to –0.11).

Conclusions: This study provides novel findings and a rigorous comparison in understanding the effects of IVR on public stigma. The findings showed that IVR and text with audio narration performed similarly and significantly in stigma reduction. Sense of embodiment and story transportation were found to be sequentially associated with public stigma reduction.

Trial Registration: Centre for Clinical Research and Biostatistics Clinical Trial Registry CUHK_Ccrb00638; https://www2.ccrb.cuhk.edu.hk/registry/public/632

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e23683) doi:10.2196/23683

KEYWORDS

immersive virtual reality; narrative persuasion; public stigma; mental health stigma; sense of embodiment; story transportation; stigma intervention

Introduction

Background

One of the biggest challenges that people with mental illness face is psychiatric stigma [1,2]. Stigma refers to a phenomenon in which people are socially discredited as a result of their social identity, health condition, or other characteristics deemed "undesirable" by the dominant group [3]. The public's stereotypical and negative beliefs, prejudicial emotional reactions, and discriminatory responses toward minority or disadvantaged groups in the society are referred to as public stigma [2,4,5]. People endorsing public stigma toward people with mental illness often make negative but inaccurate assumptions about them, including that people with mental illness are violent, dangerous, weak, and childlike [2,6-10]. People with public stigma also tend to negatively label people with mental illness as "them" and regard them as fundamentally different and undesirable to be part of "us" [2,11].

Public stigma often hampers the recovery and reintegration of people with mental illness into society [9,12]. For example, the public who hold prejudice toward people with mental illness as childlike and having mental illness due to weak character may discriminate against them through supporting coercive treatment [13]. The public who find people with mental illness fearful may call for institutional segregation and avoidance [13]. People with mental illness are often deprived of employment opportunities [5,10,14] or are falsely alleged to be involved in violent crimes [2]. Recognizing the existence of public stigma, some people with mental illness may concur and internalize these public views, leading to self-stigma [2,10,15,16]. The "why try" model has illustrated the complex process of how people with mental illness may be aware of, agree with, and apply the stereotypes of mental illness [17]. As a result, self-esteem and self-efficacy are undermined, which yields "why try" responses in pursuit of life goals [17]. People with mental illness may think they are not worthy of attaining life achievements [17] and have reduced help-seeking behavior [1,5,14,16,17]. Regarding these alarming consequences brought by stigma, varied efforts have been advocated to reduce public stigma and empower people with mental illness.

Prior Work of Combating Public Stigma

Past efforts in combating public stigma found education and contact between the public and people with mental illness to be the most effective strategies [5,7,18]. Education mainly aims at replacing stigmatizing myths with accurate information and knowledge, and contact attempts to replace negative stereotypes with a positive experience through interpersonal contact on an equal footing [7]. Education and contact significantly led to positive attitudinal change. Among all, contact seemed to yield the most efficacious effect in reducing public stigma, especially for adults [2,5,19].

However, as in vivo contact required presenters to travel to venues and be well trained, it could be tedious and hard to implement widely and frequently [4]. Researchers have therefore investigated the possibilities of using other types of contact [4,20,21]. Filmed contact has been used, and it yielded similar or weaker effects than in vivo contact [4,19-21]. Recently, researchers have begun to investigate the effect of immersive virtual reality (IVR) in reducing public stigma, although this effort is limited [22].

Reducing Public Stigma With Immersive Virtual Reality

With technological advancement, IVR becomes a possible option to reduce negative stereotypes and prejudice. IVR allows people to be put into another character by changing their body representation [23-25]. It is done by a process referred to as *virtual embodiment*, in which individuals can see a virtual body substituting their bodies through a head-tracked head-mounted display [23,24]. Some can even wear body-tracking suits to induce synchronous virtual body movements [24]. Previous research revealed that IVR could successfully induce ownership over bodies and identities of outgroups and significantly reduce negative stereotypes, including reducing implicit racial bias [24] and stigma toward older adults [26]. The public generally regard people with mental illness as different and outgroup members [2,11]. With IVR, participants could have first-person experiences as people with mental illness. They might regard people with mental illness as ingroup members and have lower stigma. With the aids of immersive experience, participants' identities and perspectives might also be changed more easily and substantially than nonimmersive means. Therefore, this study applies IVR and compares it to a nonimmersive medium to investigate its effects in reducing public stigma toward mental illness.

In previous IVR research related to mental health public stigma, they mainly focused on the simulation of schizophrenic symptoms [22,27]. Instead of reducing stigma, simulation of symptoms through IVR further increased social distance with people with schizophrenia [22]. Simulation of auditory hallucination without using IVR also yielded similar results [27-29]. In these studies, some comparison groups even reduced greater stigma than the simulation conditions [22,29]. Often, the comparison groups highlighted other components than just symptoms, such as feelings, treatments, difficulties, and accomplishments of a person with mental illness [22,29]. Similarly, in the few virtual reality studies, both immersive and nonimmersive, that had successfully reduced public stigma, the simulation condition was either paired with written empathetic tasks [22] or participants could get information more than symptoms [30]. For example, participants could understand symptoms, biographies of people with mental illness, and false negative beliefs in the virtual reality [30]. They were found to

XSL•FO
RenderX

have decreased stereotypes and perceived dangerousness of people with schizophrenia [30].

Therefore, rather than emphasizing symptom simulation, this study took a person-centered approach by simulating the daily life interactions and stigmatized experience of a person with mental illness. We hypothesized that IVR that immersed participants into the everyday living of someone with mental illness could significantly reduce public stigma toward mental illness.

Hypothesized Mediation Models of Immersive Virtual Reality and Public Stigma

Sense of Embodiment as Mediator

To examine how the IVR might affect public stigma, mediation analyses were also conducted. To expand the proposed idea of IVR changing stigma through virtual embodiment, one proposed mediator was sense of embodiment. It refers to the sensations of being inside, controlling, and having ownership over an artificial body [23]. It is possible because our body schema and the concept of the self are labile [31]. Through mapping the physical body to the virtual body, our body schema can be radically altered to the virtual one [32]. As an avatar's body can embed different social roles or meanings, our social identities may also be altered accordingly [33]. Previous IVR literature on racial bias had suggested that with sense of embodiment, participants embedded new physical and conceptual identities [33]. Other racial groups might become in-group to the participants, leading to generalization of affective processing and positive evaluation of the race and a reduction in racial bias [33]. Thus, this study hypothesized that sense of embodiment could mediate IVR's relationship with stigma reduction though changing body schema and social identity.

Story Transportation as Mediator

Another proposed mediator was story transportation. It describes how much an individual can be immersed in a story [34,35]. With strong story transportation, readers can have their cognition, affection, and attention dominated by the story [34-37]. Story transportation was suggested to have various effects on belief and stigma change [38-43]. It might reduce cognitive and elaborative activities that resist persuasion [42,43]. It could also elicit broad affective responses that allow participants to be more receptive to belief change [42]. People might also have more vivid mental images that make contents and messages more memorable, leading to a greater intention of belief change [42,43]. A previous empirical study had shown that with story transportation, participants felt related to the story protagonist through reading a written vignette and had reduced social distance with people with depression [37]. With various special features of IVR, such as the first-person field of view and tracking of head movements, IVR might result in stronger engagement with the story [44], thus stronger stigma change than a nonimmersive medium.

Sense of Embodiment and Story Transportation as Sequential Mediators

Besides testing a sense of embodiment and story transportation as stand-alone mediators, this study also attempted to explore the possible sequential effects of sense of embodiment and story transportation in the same mediation model. In this study, participants would be experiencing the story as the protagonist. Therefore, embodying the avatar could readily enable the participants to become the story protagonist themselves and be part of the virtual settings. Thus, it was hypothesized that sense of embodiment might increase story transportation, leading to reduced public stigma.

Methods

Trial Registration

This study was approved by the Survey and Behavioral Research Ethics Committee at The Chinese University of Hong Kong (approval reference number: SBRE-18-078). It is registered at The Centre for Clinical Research and Biostatistics Clinical Trial Registry (trial registration number: CUHK_CCRB00638).

Recruitment

The study was conducted from January to March 2019. A single-blinded randomized controlled trial was used for this study. According to previous studies that aimed at changing attitudes to outgroup members using IVR, a medium effect size was observed [24]. Meta-analysis showed that video contact and in vivo contact had medium to large effect sizes respectively on changing stigmatizing attitude [19]. With IVR having stronger perceptive stimulation than video contact, but less intense and interactive than in vivo contact, a medium effect size was hypothesized. Assuming a medium effect size (η_p^2=0.06) with 80% power and a probability of a type I error of 0.05, a total sample size of 141 was needed. Participants were recruited through the mass mail system of The Chinese University of Hong Kong. The purpose of this study and assignment of conditions were concealed until the debriefing session at the end of the entire study. The email and informed consent form only stated that the experiment would be related to IVR. Since the content of the interventions were in Chinese, participants who were 18 years or older and fluent in Chinese were eligible for this study. After completing the pre-experiment questionnaire online, participants who failed to answer attention check questions correctly (eg, "Please choose extremely disagree.") were excluded to ensure that participants were answering the questions diligently. Participants who failed to provide correct contact information or any available experiment time slots were also excluded. Afterward, each participant was randomly assigned into one of the three conditions: IVR animation (immersive animation), the first-person audio narration of the same story (text), and IVR space control (control). Simple randomization was used with computerized random numbers. Upon signing the consent form online with checkboxes, participants were administered their assigned condition and completed an online postexperiment questionnaire immediately in the experiment room and 1 week after the experiment at home. A total of HK $50 (US $6.44) were given to each participant as a remuneration of time after the accomplishment of the experiment. Responses from participants who completed the postexperiment and follow-up questionnaires were analyzed.

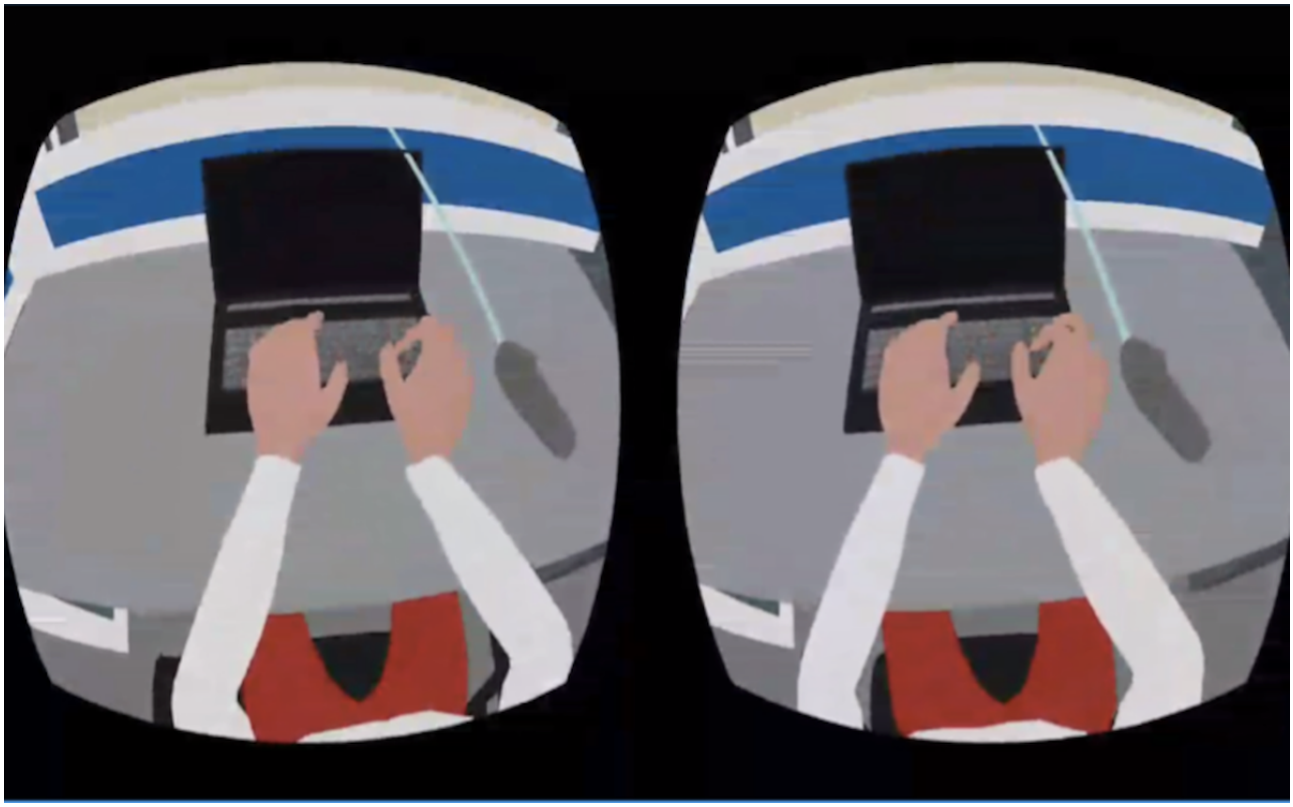
XSL•FO
RenderX

Experimental Conditions

All experimental conditions were around 10 minutes long. To avoid any placebo effect and to keep other factors constant, all three conditions were conducted with a researcher-assisted virtual reality headset, Oculus Go (Facebook), which could provide 360° head-tracking IVR with corresponding audio and sound effects. With the exception of the researcher assisting with the operation of the IVR headset at the commencement of the experiment, all presentation of experimental content was automatic with minimal technical support from the researcher to minimize researcher bias. All participants received their assigned condition in a dark room while seated on a swivel chair.

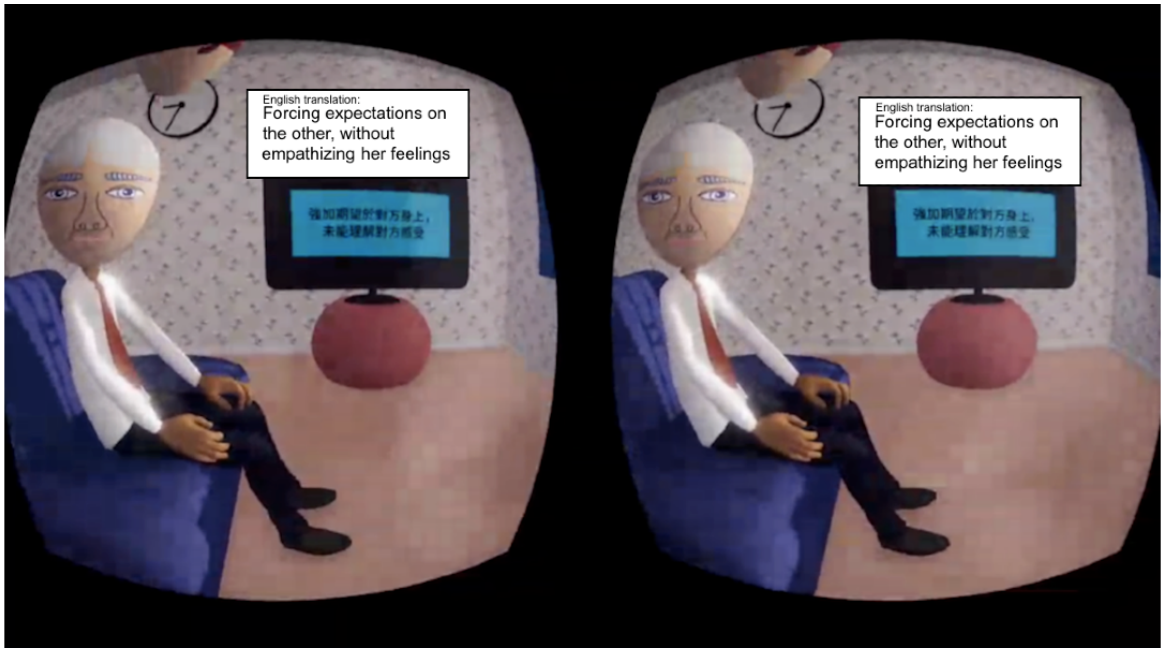
In the immersive animation condition, participants were immersed in an animated story as the female protagonist, Yan, who had mixed anxiety and depressive disorder. In the IVR, Yan's body would be in the position of the participants' bodies. Participants would be seeing things from Yan's perspective (Figure 1). The story started with Yan working at the office. The scene then changed to Yan sitting on a sofa alone at home, where stigma from colleagues and pressure from work and family were illustrated. Afterward, Yan talked to her uncle who was visiting. The uncle was an antagonist who stigmatized and imposed more pressure onto Yan. Yan's field of view would sink while the uncle talked, and everything else would seem larger and higher to create a sense of inferiority. There were also pop-up messages on the side to illustrate the problematic issues in the uncle's speech during the conversation (Figure 2). Examples included trivializing experience, attributing all responsibilities to the person, and advice-giving without consideration of appropriateness.

Figure 1. Screenshot of participants seeing the female protagonist's, Yan's, body in the immersive animation condition.



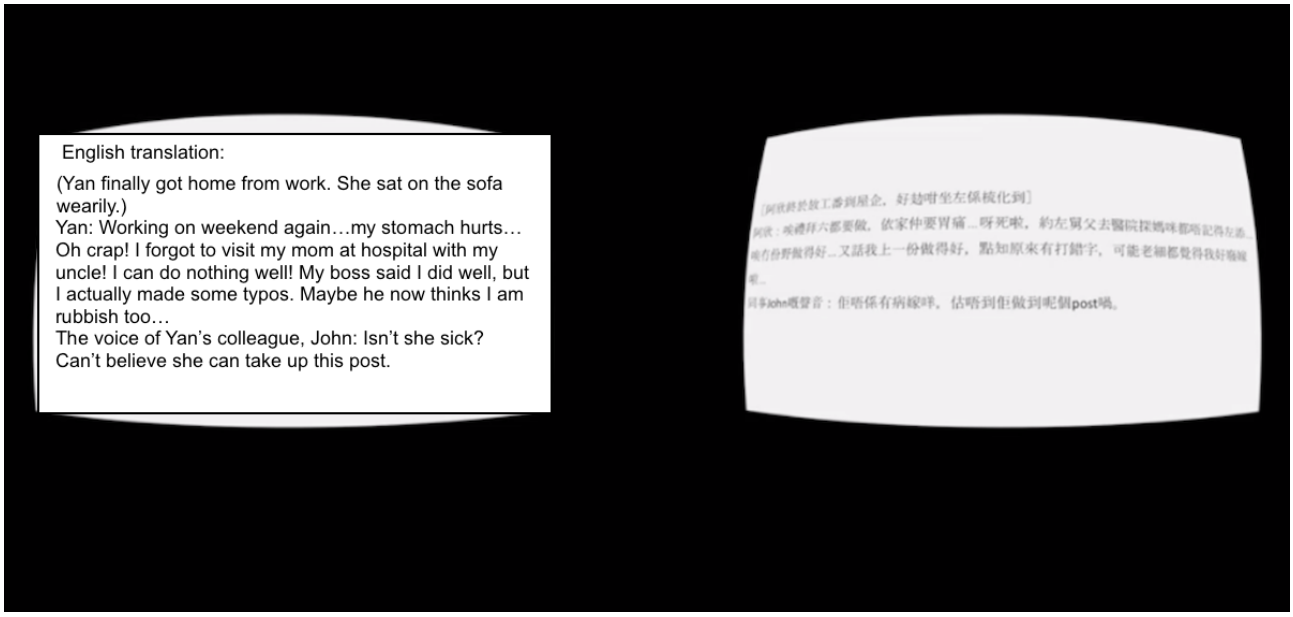
XSL•FO
RenderX

Figure 2. Screenshot of Yan talking to her uncle, and showing how messages pop up to illustrate problematic issues in the uncle's speech in the immersive animation condition.



In the text condition, participants read the text and listened with audio narration to the same story as the immersive animation condition. Participants still wore the virtual reality headset to avoid any placebo effect. However, they read the story in a textual and 2D format on a white background with black text (Figure 3). To investigate the separate effects of visual inputs in the IVR condition, the first-person voice-over and sound effects in the IVR remained unchanged here and accompanied the text in accordance to the story development. This condition was a conventional medium that provided a rigorous comparison to the immersive animation condition. It kept the story components identical except for the visual inputs and immersive experience in the IVR. This condition was compared with the immersive animation condition in both the main effects of stigma reduction and mediation analyses. It allowed examination of possible mechanisms of how IVR might affect public stigma differently than nonimmersive means through proposed mediators.

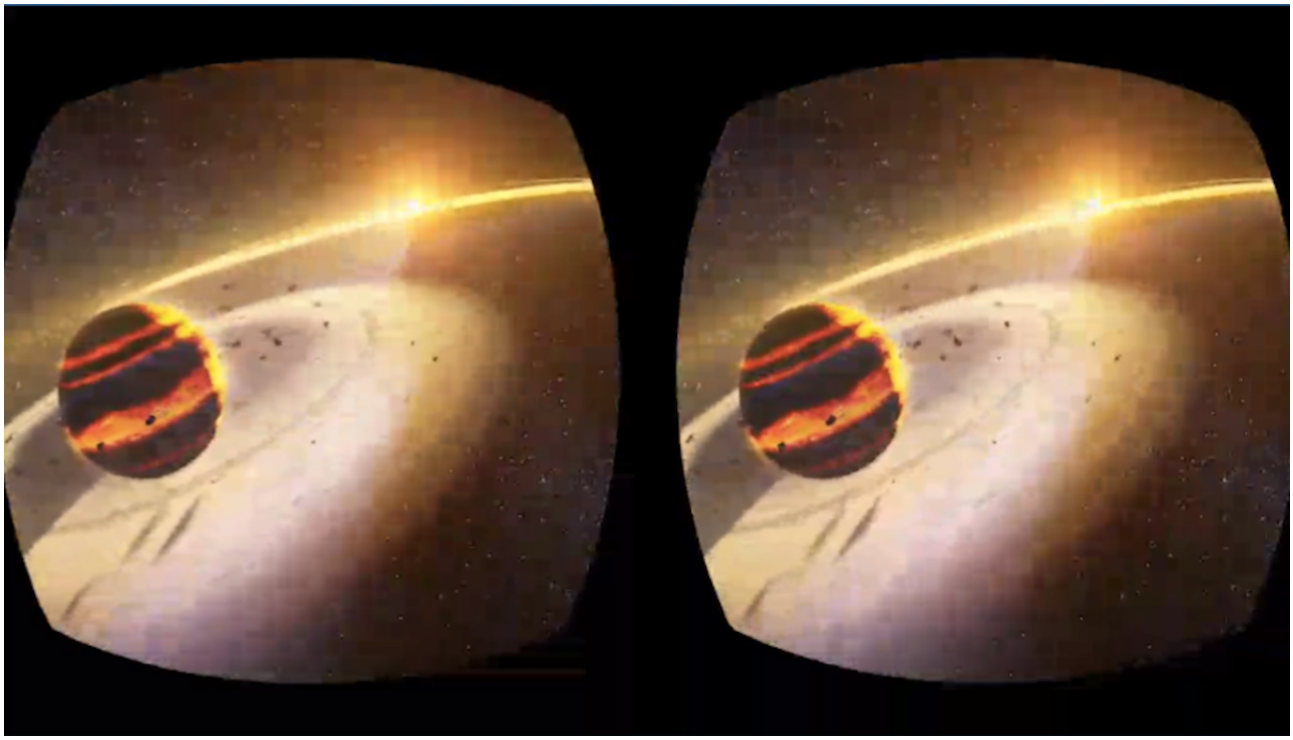
Figure 3. Screenshot of the text condition.



In the control condition, participants watched a 360° exoplanet virtual reality video with corresponding audio and sound effects. The exoplanet video was titled "Take a Virtual Reality tour of six REAL exoplanets" on YouTube with its copyright owned by the University of Exeter, 2018 (Figure 4). The video was created by the University of Exeter Astrophysics group in partnership with We the Curious and Engine House VFX. This video was used in the study by permission of the University of Exeter.

XSL•FO
RenderX

Figure 4. Screenshot of the space control condition.



Measures

The Main Effect on Public Stigma

To measure changes in public stigma, the 21-item Public Stigma and Acceptance Scale [45] was used at pre-experiment, postexperiment, and 1-week follow-up. It measures public stigma and acceptance toward mental illness on a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree). It consists of two subscales, 12 items measuring public stigma (eg, "People with mental illness are a burden to the society.") and 9 items measuring personal advocacy toward accepting people with mental illness (eg, "I take the initiative to reach out to people with mental illness."). The higher the mean scores of the public stigma subscale and the lower the mean scores of the personal advocacy subscale, the greater the stigma and the lower the acceptance toward mental illness. For the total mean scores, items in the personal advocacy subscale were reverse coded so that higher total mean scores indicated greater stigmatizing attitudes. In this study, total mean scores of all items were used to measure public stigma. Possible total mean scores ranged from 1 to 6. The Cronbach α of the public stigma and personal advocacy subscales were .92 and .87, respectively. The overall Cronbach α was .93 for the entire scale in this study.

Mediators

Sense of Embodiment

To account for possible mediation effects, sense of embodiment and story transportation were measured post experiment in the immersive animation and text conditions. Since the topic of interest was on what might influence change in public stigma and how IVR's effect differed from the conventional audio narration, scales on mediation were not administered in the control condition.

Although previous studies have measured sense of embodiment [24,46,47], the items were mostly context-specific to the virtual environments and avatars in those studies. Therefore, with no standardized scale available, a 7-item Sense of Embodiment Scale was developed to fit this experiment context based on previous virtual reality studies [24,46,47] and rubber hand illusion literature [48]. Seven items were developed based on components that were commonly shared among the previous literature, such as ownership (eg, "I felt like I was looking at my own body when I looked at Yan's.") and agency (eg, "If I had wanted, I felt like I could have controlled Yan's body."). Items were rated on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree; see Textbox 1 for complete scale items). The higher mean score indicated a greater sense of embodiment. The Sense of Embodiment Scale had a Cronbach α of .85 in this study.

XSL•FO
RenderX

Textbox 1. Items in the Sense of Embodiment Scale, with items 6 and 7 as reversed items.

1. I felt as if I was looking at my own body when I looked at Yan's.
2. When Yan was sitting on the sofa, I felt as if I was sitting on the sofa.
3. I felt as if Yan's body was my body.
4. When the uncle was talking to Yan, I felt as if he was talking to me.
5. If I had wanted, I felt like I could have controlled Yan's body.
6. I found it difficult to get into the role of Yan.
7. I felt as if I was an outsider to the story.

Story Transportation

The Transportation Scale [36] measures how well an individual is absorbed into a story. Some wordings of the 15-item scale were modified to fit the current context (eg, changing "While reading the narrative I had a vivid image of Katie" to "While experiencing the story I had a vivid image of Yan."). The scale was rated on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). The higher mean score indicated greater devotion in the story. The Transportation Scale had a Cronbach α of .79 in this study.

The Public Stigma and Acceptance Scale and the Sense of Embodiment Scale were originally developed in Chinese. The Transportation Scale was translated into Chinese and back-translated into English to check on conceptual equivalence.

Data Analysis

Analyses were conducted using SPSS 24.0 for Mac (IBM Corp). First, descriptive statistics were obtained. Information related to age, gender, education, occupation, sexual orientation, religious belief, marital status, mental illness history, and previous social contact with people with mixed anxiety and depressive disorder was obtained. Chi-square tests on nominal demographic variables and one-way analysis of variance (ANOVA) on continuous variables were conducted to examine any systematic differences on demographics among the 3 groups. If there were any systematic differences among the groups, results would be adjusted, controlling for possible confounders. Second, to compare the change in public stigma, repeated measures ANOVA was used. The main effect of groups (immersive animation vs text vs control), time (pre-experiment, postexperiment, follow-up), and their interaction effect (time × group) were examined. The level of statistical significance was set at a $P \leq .05$ threshold (two-tailed). If a significant effect was found, post hoc analysis was conducted. Bonferroni-adjusted pairwise comparisons were used to avoid inflated type I error. Third, to test the proposed mediation models, Pearson correlations of condition (immersive animation vs text), sense of embodiment, story transportation, and follow-up public stigma were obtained. A simple mediation analysis was conducted with the condition (immersive animation vs text), sense of embodiment, and follow-up public stigma. A similar simple mediation analysis was also conducted using story transportation as the mediator. To further investigate the relationship of sense of embodiment and story transportation, sequential mediation analysis was also conducted. The condition (immersive animation vs text) and follow-up public stigma were put into the model, having sense of embodiment and story transportation as sequential mediators. The level of statistical significance was set at $P \leq .05$ threshold (two-tailed) and determined by 95% CIs. PROCESS on SPSS was used to generate all the mediation analyses. Model 4 was used to generate simple mediation analyses, and model 6 was used to generate sequential mediation analysis according to the suggestion of Hayes [49].

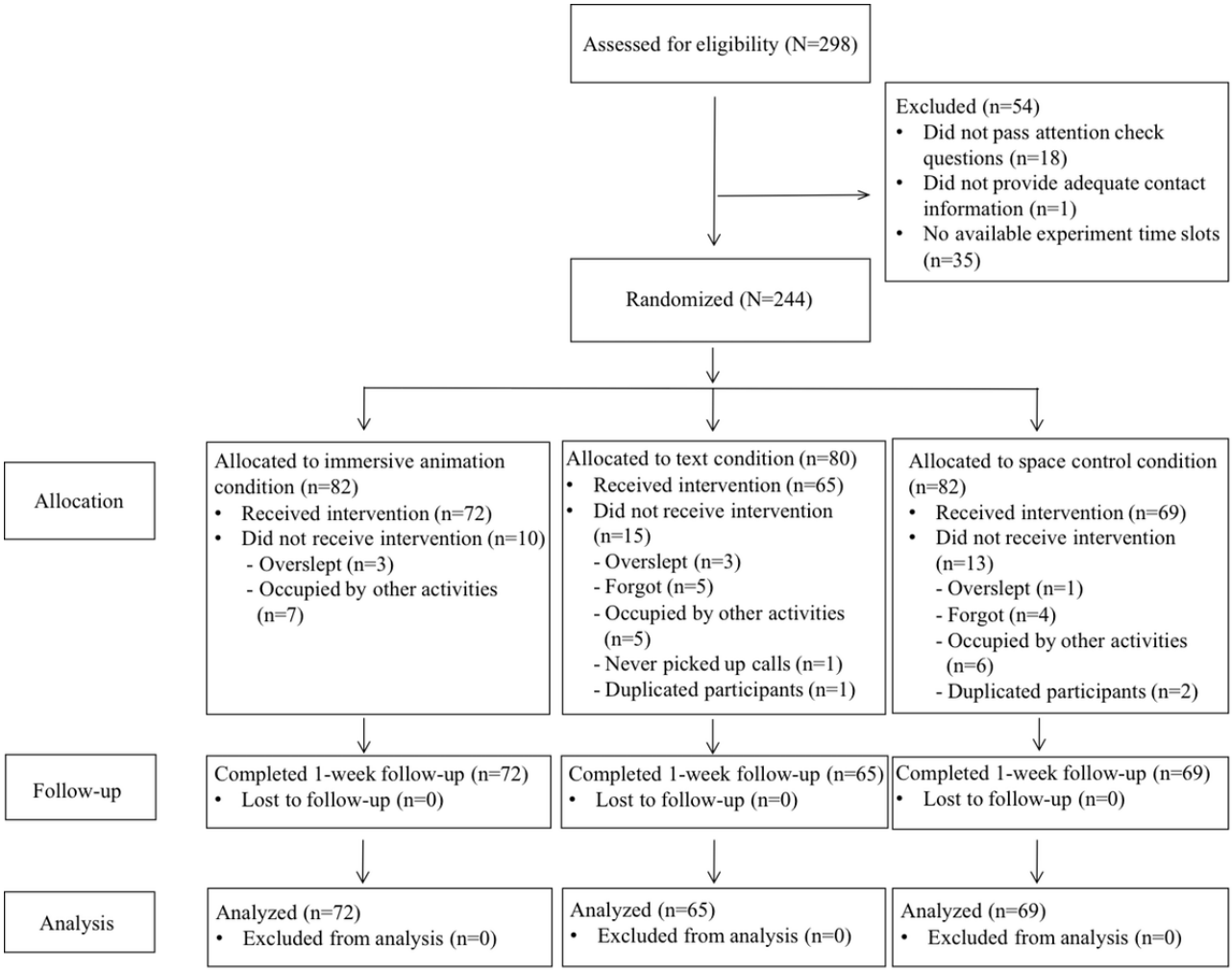
Results

Participants

Data from 206 participants were analyzed. Out of the 298 participants who provided informed consent and completed the pre-experiment questionnaire, 18 participants who failed to answer the attention check questions correctly, 1 participant who failed to provide adequate contact information, and 35 participants who failed to provide available experiment time slots were excluded (Figure 5). A total of 244 participants were randomized into the three conditions. A total of 35 participants who did not show up and 3 participants who applied to the same experiment twice were excluded. A total of 206 participants went through the experiment and completed questionnaires post experiment and at a 1-week follow-up. The final sample size was 72 for the immersive animation condition, 65 for the text condition, and 69 for the control condition, with a total sample size of 206.

XSL•FO
RenderX

Figure 5. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant recruitment.



For demographic information, see Table 1. The mean age of participants was 21.76 (SD 5.04, range 18-64) years. Out of the 206 participants, 114 (55.3%) were females and 91 (44.2%) were males. Most participants had a Bachelor's degree (n=166, 80.6%) and were single (n=198, 96.1%), students (n=173, 84.0%), and heterosexual (n=183, 88.8%). Most of them had no mental illness history (n=198, 96.1%) and no religious belief (n=153, 74.3%). Among the 206 participants, 92 (44.7%) had social contact with people with mixed anxiety and depressive disorder before. Chi-square tests on demographic variables and one-way ANOVA on continuous variables were conducted. No systematic differences in any demographic variables among the three conditions were found. Demographics variables were not expected to impact the results in this study.

XSL•FO
RenderX

Table 1. Participant demographic data.[a]

Characteristic	Total (N=206)	Immersive animation (n=72)	Text (n=65)	Control (n=69)	*P* value difference across conditions
Age (years), mean (SD)	21.76 (5.04)	21.57 (4.18)	21.95 (6.31)	21.77 (4.55)	.91
Gender, n (%)					.68
Female	114 (55.3)	38 (52.8)	35 (53.8)	41 (59.4)	
Male	91 (44.2)	33 (45.8)	30 (46.2)	28 (40.6)	
Others	1 (0.5)	1 (1.4)	0 (0.0)	0 (0.0)	
Education, n (%)					.37
High school	6 (2.9)	1 (1.4)	1 (1.5)	4 (5.8)	
Bachelor's degree	166 (80.6)	60 (83.3)	54 (83.1)	52 (75.4)	
Master's degree	27 (13.1)	10 (13.9)	7 (10.8)	10 (14.5)	
PhD	2 (1.0)	1 (1.4)	0 (0.0)	1 (1.4)	
Others	5 (2.4)	0 (0.0)	3 (4.6)	2 (2.9)	
Faculty, n (%)					.93
Arts	25 (12.1)	9 (12.5)	8 (12.3)	8 (11.6)	
Business	51 (24.8)	17 (23.6)	17 (26.2)	17 (24.6)	
Education	14 (6.8)	7 (9.7)	4 (6.2)	3 (4.3)	
Engineering	16 (7.8)	5 (6.9)	5 (7.7)	6 (8.7)	
Law	2 (1.0)	0 (0.0)	1 (1.5)	1 (1.4)	
Medicine	25 (12.1)	5 (6.9)	11 (16.9)	9 (13.0)	
Science	27 (13.1)	12 (16.7)	6 (9.2)	9 (13.0)	
Social science	44 (21.4)	16 (22.2)	12 (18.5)	16 (23.2)	
Others	2 (1.0)	1 (1.4)	1 (1.5)	0 (0.0)	
Occupation, n (%)					.29
Student	173 (84.0)	63 (87.5)	57 (87.7)	53 (76.8)	
Full-time employment	29 (14.1)	8 (11.1)	7 (10.8)	14 (20.3)	
Others	4 (1.9)	1 (1.4)	1 (1.5)	2 (2.9)	
Sexual orientation, n (%)					.68
Heterosexual	183 (88.8)	66 (91.7)	58 (89.2)	59 (85.5)	
Homosexual	11 (5.3)	3 (4.2)	4 (6.2)	4 (5.8)	
Bisexual	7 (3.4)	2 (2.8)	2 (3.1)	3 (4.3)	
Others	5 (2.4)	1 (1.4)	1 (1.5)	3 (4.3)	
Religious belief, n (%)					.15
No religious belief	153 (74.3)	47 (65.3)	53 (81.5)	53 (76.8)	
Christian	43 (20.9)	22 (30.6)	9 (13.8)	12 (17.4)	
Others	10 (4.9)	3 (4.2)	3 (4.6)	4 (5.8)	
Marital status, n (%)					.42
Single	198 (96.1)	68 (94.4)	62 (95.4)	68 (98.6)	
Married	8 (3.9)	4 (5.6)	3 (4.6)	1 (1.4)	
Mental illness history, n (%)					.46
Yes	8 (3.9)	4 (5.6)	1 (1.5)	3 (4.3)	
No	198 (96.1)	68 (94.4)	64 (98.5)	66 (95.7)	

XSL•FO
RenderX

Characteristic	Total (N=206)	Immersive animation (n=72)	Text (n=65)	Control (n=69)	*P* value difference across conditions
Social contact with people with MADD[b], n (%)					.64
Yes	92 (44.7)	29 (40.3)	31 (47.7)	32 (46.4)	
No	114 (55.3)	43 (59.7)	34 (52.3)	37 (53.6)	

[a]All percentages may not total to 100% due to rounding.

[b]MADD: mixed anxiety and depressive disorder.

Changes in Public Stigma

Repeated measures ANOVA on public stigma yielded a significant interaction (time × group) effect ($F_{4,404}$=5.2; *P*<.001; η_p^2=0.05). The effect size was small. Bonferroni-adjusted pairwise comparisons showed that the immersive animation (mean 2.73, SD 0.64), text (mean 2.78, SD 0.68), and control (mean 2.84, SD 0.64) conditions did not differ in public stigma pre-experiment (*P*>.99). Post experiment, the immersive animation group (mean 2.45, SD 0.63) had significantly lower public stigma than the control group (mean 2.80, SD 0.63; *P*=.003). The text group (mean 2.47, SD 0.62) yielded similar results by having significantly lower public stigma than the control group (*P*=.007). A similar pattern of results was also observed at the 1-week follow-up. The immersive animation group (mean 2.52, SD 0.66) had significantly lower public stigma than the control group (mean 2.82, SD 0.61; *P*=.02). The text group (mean 2.53, SD 0.68) also had significantly lower public stigma than the control group at follow-up (*P*=.03). However, the immersive animation and text groups did not differ from each other significantly at postexperiment and at the 1-week follow-up (*P*>.99; see Table 2 for descriptive statistics and Table 3 for detailed pairwise comparison).

Table 2. Public stigma scores across conditions at different time points.

Time points	Immersive animation (n=72), mean (SD)	Text (n=65), mean (SD)	Control (n=69), mean (SD)
Pre-experiment	2.73 (0.64)	2.78 (0.68)	2.84 (0.64)
Postexperiment	2.45 (0.63)	2.47 (0.62)	2.80 (0.63)
Follow-up	2.52 (0.66)	2.53 (0.68)	2.82 (0.61)

Table 3. Detailed Bonferroni-adjusted pairwise comparison of public stigma between conditions at different time points.

Time points	Condition	Condition for comparison	Mean difference (SE)	*P* value
Pre-experiment	Immersive animation	Text	–0.05 (0.11)	>.99
Pre-experiment	Immersive animation	Control	–0.11 (0.11)	.99
Pre-experiment	Text	Control	–0.06 (0.11)	>.99
Postexperiment	Immersive animation	Text	–0.02 (0.11)	>.99
Postexperiment	Immersive animation	Control	–0.36 (0.11)	.003
Postexperiment	Text	Control	–0.33 (0.11)	.007
Follow-up	Immersive animation	Text	–0.01 (0.11)	>.99
Follow-up	Immersive animation	Control	–0.30 (0.11)	.02
Follow-up	Text	Control	–0.29 (0.11)	.03

In the immersive animation group, a significant time effect was found ($F_{2,202}$=20.0; *P*<.001; η_p^2=0.165). The effect size was large. The immersive animation group had significantly lower public stigma at postexperiment (mean 2.45, SD 0.63) and follow-up (mean 2.52, SD 0.66) than pre-experiment (mean 2.73, SD 0.64; both with *P*<.001). Public stigma did not significantly differ between postexperiment and follow-up (*P*=.19), implying that the effect at postexperiment was sustained at follow-up for the immersive animation group. In the text group, a significant time effect was also found ($F_{2,202}$=21.3; *P*<.001; η_p^2=0.174). The effect size was large. The text group followed a similar pattern of time effect as the immersive animation group. The text group had significantly lower public stigma at postexperiment (mean 2.47, SD 0.62) and follow-up (mean 2.53, SD 0.68) than pre-experiment (mean 2.78, SD 0.68; both with *P*<.001). No significant difference between postexperiment and follow-up was found for the text group (*P*=.51). In the control group, no significant time effect was found ($F_{2,202}$=0.37; *P*=.69; η_p^2=0.004), and the mean score was 2.84 (SD 0.64) at pre-experiment, 2.80 (SD 0.63) at postexperiment, and 2.82 (SD 0.61) at follow-up.

XSL•FO
RenderX

Mediation Analyses

The mean scores of sense of embodiment, story transportation, and follow-up public stigma are reported in Table 4. Pearson correlations of variables included in the mediation analyses, including condition (immersive animation vs text), sense of embodiment, story transportation, and follow-up public stigma are also reported in Table 4.

Table 4. Correlations and descriptive statistics of variables included in the mediation analyses.

Variables	Correlations of variables included in the mediation analyses			
	Condition (immersive animation vs text)	Sense of embodiment	Story transportation	Follow-up public stigma
Condition (immersive animation vs text)[a]				
r	1	0.49	0.17	–0.004
P value	—[b]	<.001	.051	>.99
Sense of embodiment				
r	0.49	1	0.512	–0.091
P value	<.001	—	<.001	.29
Story transportation				
r	0.17	0.512	1	–0.354
P value	.051	<.001	—	<.001
Follow-up public stigma				
r	–0.004	–0.091	–0.354	1
P value	>.99	.29	<.001	—
Overall, n	137	137	137	137
Overall, mean (SD)	—[c]	4.25 (1.10)	5.00 (0.66)	2.52 (0.67)
Immersive animation condition, mean (SD)	—[c]	4.75 (0.95)	5.11 (0.59)	2.52 (0.66)
Text condition, mean (SD)	—[c]	3.69 (0.99)	4.89 (0.71)	2.53 (0.68)

[a]Condition was dummy coded with text condition as 0 and immersive animation as 1 in the correlation and mediation analysis. Control condition was not included in the mediation analysis.

[b]Not applicable.

[c]Condition was a dichotomous variable. Mean and SD were not applicable.

Consistent with the correlation results (Table 4), the simple mediation analysis showed that sense of embodiment was significantly associated with condition (β=1.07; t_{135}=6.45; *P*<.001; Figure 6). However, sense of embodiment was not significantly associated with public stigma (β=–.07; t_{135}=–1.19; *P*=.23). Therefore, the indirect effect was not significant (95% CI –0.22 to 0.46). The direct effect from condition to public stigma was also not significant (95% CI –0.18 to 0.33).

For story transportation, the result was also consistent with the correlation analysis (Table 4). The simple mediation analysis showed that the condition was not significantly associated with story transportation (β=.22; t_{135}=1.97; *P*=.051; Figure 7). However, story transportation was significantly and negatively associated with public stigma (β=–.37; t_{135}=–4.44; *P*<.001). The indirect effect of condition to story transportation to public stigma was not significant (95% CI –0.18 to 0.00). The direct effect from condition to public stigma was also not significant (95% CI –0.14 to 0.29).

XSL•FO
RenderX

Figure 6. Simple mediation model of the condition, sense of embodiment, and public stigma.

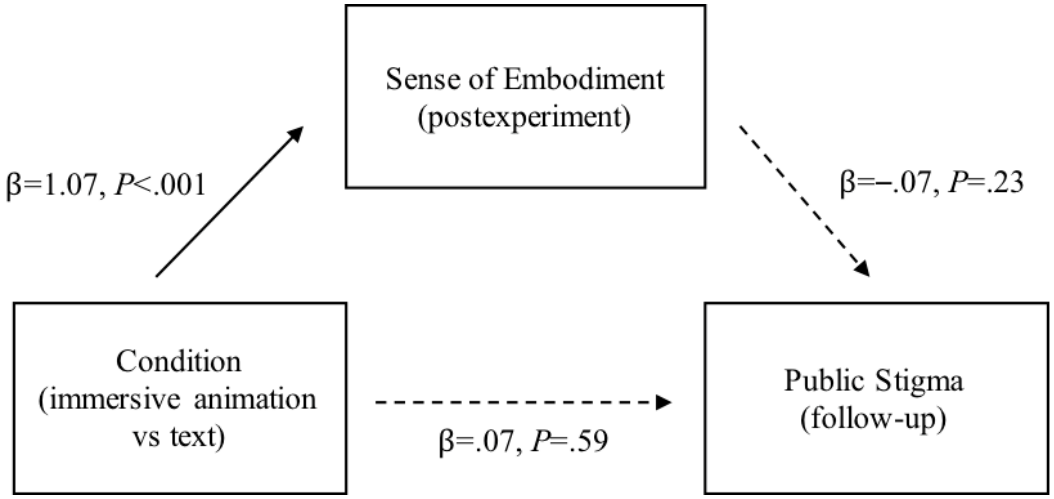
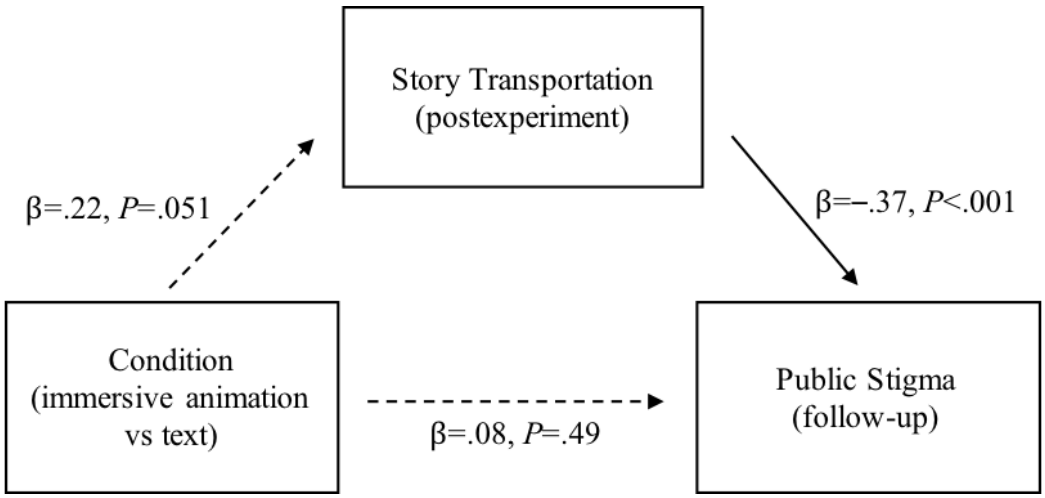


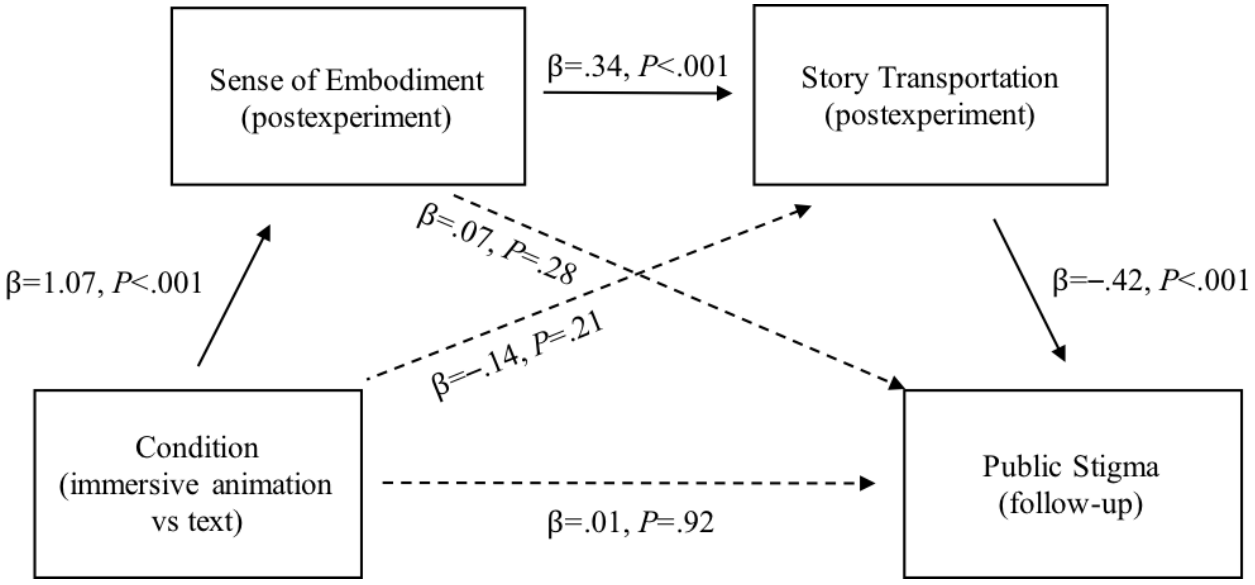
Figure 7. Simple mediation model of the condition, story transportation, and public stigma.



Sense of embodiment and story transportation were then put into the same sequential mediation analysis to investigate their relationship on follow-up public stigma. From the sequential mediation analysis (Figure 8), the relationship of conditions and follow-up public stigma was significantly and sequentially mediated by sense of embodiment and story transportation as hypothesized. In this sequential mediation model, the condition was significantly associated with sense of embodiment (β=1.07; t_{134}=6.45; *P*=<.001) but not directly with story transportation (β=–.14; t_{134}=–1.26; *P*=.21). Sense of embodiment was significantly associated with story transportation (β=.34; t_{134}=6.68; *P*<.001) but was not significantly associated with public stigma (β=.07; t_{134}=1.09; *P*=.28). For story transportation, it was significantly and negatively associated with public stigma (β=–.42; t_{134}=–4.40; *P*<.001). The indirect effect of condition to sense of embodiment then to story transportation and finally to follow-up public stigma was significant (Figure 8). The standardized indirect effect was –0.15. The bootstrapped unstandardized indirect effect was also –0.15 (95% CI –0.38 to –0.11). However, the direct effect from condition to follow-up public stigma was not significant (95% CI –0.23 to 0.26).

XSL•FO
RenderX

Figure 8. Sequential mediation model of the condition, sense of embodiment, story transportation, and public stigma.



Discussion

Overall Effect of Immersive Animation on Public Stigma

This study compared the effect of IVR and text with audio narration in reducing public stigma. Both IVR and text with audio narration successfully reduced public stigma, and the effect was sustained at the 1-week follow-up. In contrast, the control space video did not elicit any change in public stigma as expected. It could be inferred that the simulation of daily interactions and stigma experienced by a person with mental illness could effectively reduce public stigma. The result also supported the hypothesis that IVR could successfully reduce public stigma.

The results were consistent with previous studies that used IVR to reduce negative stereotypes toward outgroup members, including other racial groups [24] and older adults [26]. In the mental health setting, it was also consistent with virtual reality literature that focused on interacting with people with mental illness and dismissing negative beliefs [30]. However, the results in this study were not consistent with previous IVR and studies that simulated symptoms of people with schizophrenia. The simulations of hallucinations had led to increases in social distance and negative attitudes toward people with schizophrenia [22,29]. The different results from this study might be due to the different emphasis of immersive content.

This study had purposely included little simulations of symptoms in the IVR. It aimed to construct the everyday living experience of the story protagonist as a person, instead of just a combination of symptoms. Although, unlike IVR studies on racial groups or older adults, there was not a change in appearance according to an outgroup member in this study, the first-person field of view and immersion as the story protagonist with mixed anxiety and depressive disorder might already have allowed participants to take on the outgroup identity and perspective as a person with mental illness [24]. On the contrary, in previous literature on stigma change toward people with schizophrenia, the simulation of a hallucination might have emphasized the existence of positive symptoms, instead of allowing participants to understand and feel other aspects of the living experience. Positive symptoms that could be perceived as uncontrollable and dangerous by the public might deter participants from engaging with people with schizophrenia [50]. Therefore, simulating positive symptoms might increase social distance and discrimination instead.

On the other hand, different from the hypothesis, IVR did not outperform text with audio narration in reducing public stigma significantly. In the few stigma studies that compared virtual reality with a rigorous comparison group, it was shown that only virtual reality with interactive components such as allowing participants to move closer to other virtual characters had significantly reduced greater stigma than the written vignettes of people with mental illness [51]. Immersive experience without active interaction did not outperform the written vignettes in stigma reduction [22]. Therefore, the visual immersive experience of IVR might not be strong enough to induce a significant difference than text with audio narration in stigma reduction. Interactive features in IVR, such as making choices or controlling objects in the virtual environment [52], or even having body or arm tracking for synchronized movements with avatars, might induce a stronger effect in stigma reduction than text. These features of IVR should be tested in the future.

Another possible explanation of the similar performance of the IVR and text conditions might be due to the type of graphics used. This IVR used animated cartoons. In the text condition, although without pictorial inputs, voice-over and detailed descriptions of the story, including the virtual environment, characters, and actions, were included. It might help with imagining the story in participants' minds. Those imaginations might perform similarly with animated cartoons in engaging the participants. Thus, having similar effects on stigma reduction. With more realistic animation, IVR might allow participants to resonate better and be more involved in the story [52]. Therefore, IVR with more realistic graphics should be investigated in the future.

XSL•FO
RenderX

Sense of Embodiment, Story Transportation, and Public Stigma

Moving on to the mediation analyses, although the condition was not directly associated with story transportation, story transportation was found to be a preceding variable directly associated with a reduction in public stigma. As mentioned previously, story transportation emphasized the cognitive and affective immersion in a story [34,36], which might elicit various cognitive and affective effects on belief change [40-43]. It might also reduce elaborative activities that resist stigma change [42,43] and elicit broad affective responses that allow participants to be more conducive to reduce their public stigma [42]. Therefore, stigma change might benefit from the broad cognitive and affective influences by story transportation.

On the other hand, although the condition was associated with sense of embodiment, sense of embodiment was not directly associated with stigma change. However, with the sequential mediation model, sense of embodiment was significantly and positively associated with story transportation, which was then linked to the reduction of public stigma. This showed that the physical sensation of being inside an avatar's body might not be sufficient for reducing public stigma. However, sense of embodiment might allow participants to feel as if they were the story protagonist and be transported in the story. Afterward, participants might have different cognitive and affective reactions from story transportation, which facilitated reduced public stigma.

The results from mediation analyses also gave more clues as to why IVR and text conditions performed similarly. The only difference between the two conditions was the absence of IVR's visual inputs and 3D immersive experience in the text condition. However, from the mediation analyses, the intense absorption of thoughts and emotions to the story was more important and closely linked with reducing public stigma than merely the physical sensation in another body. The complexity of stigma change called for more holistic simulation of experiences that allow participants to be more absorbed in the stories, lives, and feelings of a person with mental illness. Thus, the differences in visual inputs or physical experience in this IVR setting might not be sufficient to induce stronger story transportation and stigma reduction than the text condition.

Limitations and Future Directions

Although this study investigated a novel medium, IVR and its effects on public stigma, several limitations were acknowledged. First, some scales in this study, such as the Sense of Embodiment Scale, were originally developed in this study due to the lack of available scales. Further validation of the newly developed scales is needed. Second, the participants in this study were mainly college students. It might pose a challenge to the generalization of results to the general population. Therefore, replication of the study on a more diverse sample is needed to understand the effects of IVR on the public. Third, moderate to long-term effects of stigma change were unknown. Although there was a 1-week follow-up in this study, the follow-up would still be considered as measuring a rather short-term effect of the interventions [5]. Longer follow-up, such as 1 month, is needed to examine how stigma change may be sustained over time.

Fourth, actual behavioral consequences were not measured. Research showed that there could be a difference between intention and actual behavior [53-55]. Future IVR studies could consider measuring participants' gaze orientation and interpersonal distance toward a person with mental illness in the virtual environment [56]. This would require participants to walk around the experiment room with full-body tracking [56]. Fifth, compared to text, IVR could be quite novel to most people. Interventions in this study all lasted 10 minutes. Participants might struggle to get accustomed to IVR within a short period. Future studies can give participants an embodiment phase or a practice session to freely familiarize with and explore their virtual bodies and the virtual environment [24,26]. Sixth, in the text condition, there were no pictorial or animated inputs. Filmed contact or videos are of greater visual similarity with IVR than text. In vivo contact even allows participants to have real life perception and interaction with people. They may allow more rigorous comparison with IVR. However, due to the lack of available presenters and funding, in vivo contact, filmed contact, and video were not available in this study. Future studies can consider including them as comparison groups.

Notwithstanding these limitations, this study was one of the few attempts to investigate the effects of IVR on public stigma by comparing it with two rigorously designed comparison groups that attempted to rule out the separate effects of IVR and text with audio narration.

Conclusion

This study provided novel findings and a rigorous comparison in understanding the effects of IVR as a medium of public stigma reduction. The findings showed that IVR and text with audio narration both performed similarly well in reducing public stigma. Sense of embodiment and story transportation were found to be sequentially associated with public stigma reduction. This study gave a more nuanced understanding to the complexity of reducing public stigma and components in building an effective medium of stigma reduction in the future.

Acknowledgments

The IVR was developed for the Jockey Club TourHeart Project, which is sponsored by the Jockey Club Charities Trust.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 393 KB - jmir_v23i7e23683_app1.pdf]

References

1. Kosyluk KA, Al-Khouja M, Bink A, Buchholz B, Ellefson S, Fokuo K, et al. Challenging the stigma of mental illness among college students. J Adolesc Health 2016 Sep;59(3):325-331. [doi: 10.1016/j.jadohealth.2016.05.005] [Medline: 27324577]
2. Rüsch N, Angermeyer MC, Corrigan PW. Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. Eur Psychiatry 2005 Dec;20(8):529-539. [doi: 10.1016/j.eurpsy.2005.04.004] [Medline: 16171984]
3. Goffman E. Stigma: Notes on the Management of Spoiled Identity. New York: Touchstone; 1986.
4. Chan JYN, Mak WWS, Law LSC. Combining education and video-based contact to reduce stigma of mental illness: "The Same or Not the Same" anti-stigma program for secondary schools in Hong Kong. Soc Sci Med 2009 Apr;68(8):1521-1526. [doi: 10.1016/j.socscimed.2009.02.016] [Medline: 19282079]
5. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]
6. Crisp AH, Gelder MG, Rix S, Meltzer HI, Rowlands OJ. Stigmatisation of people with mental illnesses. Br J Psychiatry 2000 Jul;177:4-7. [doi: 10.1192/bjp.177.1.4] [Medline: 10945080]
7. Corrigan PW, Green A, Lundin R, Kubiak MA, Penn DL. Familiarity with and social distance from people who have serious mental illness. Psychiatr Serv 2001 Jul;52(7):953-958. [doi: 10.1176/appi.ps.52.7.953] [Medline: 11433114]
8. Link BG, Phelan JC, Bresnahan M, Stueve A, Pescosolido BA. Public conceptions of mental illness: labels, causes, dangerousness, and social distance. Am J Public Health 1999 Sep;89(9):1328-1333. [doi: 10.2105/ajph.89.9.1328] [Medline: 10474548]
9. Pinfold V, Toulmin H, Thornicroft G, Huxley P, Farmer P, Graham T. Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. Br J Psychiatry 2003 Apr;182:342-346. [doi: 10.1192/bjp.182.4.342] [Medline: 12668411]
10. Seeman N, Tang S, Brown AD, Ing A. World survey of mental illness stigma. J Affect Disord 2016 Jan 15;190:115-121. [doi: 10.1016/j.jad.2015.10.011] [Medline: 26496017]
11. Link BG, Phelan JC. Conceptualizing stigma. Annu Rev Sociol 2001 Aug;27(1):363-385. [doi: 10.1146/annurev.soc.27.1.363]
12. Wahl OF. Stigma as a barrier to recovery from mental illness. Trends Cogn Sci 2012 Jan;16(1):9-10. [doi: 10.1016/j.tics.2011.11.002] [Medline: 22153582]
13. Corrigan PW, Watson AC. Understanding the impact of stigma on people with mental illness. World Psychiatry 2002 Feb;1(1):16-20 [FREE Full text] [Medline: 16946807]
14. Wahl OF. Mental health consumers' experience of stigma. Schizophr Bull 1999;25(3):467-478. [doi: 10.1093/oxfordjournals.schbul.a033394] [Medline: 10478782]
15. Livingston JD, Boyd JE. Correlates and consequences of internalized stigma for people living with mental illness: a systematic review and meta-analysis. Soc Sci Med 2010 Dec;71(12):2150-2161. [doi: 10.1016/j.socscimed.2010.09.030] [Medline: 21051128]
16. Mak WWS, Cheung RYM. Self-stigma among concealable minorities in Hong Kong: conceptualization and unified measurement. Am J Orthopsychiatry 2010 Apr;80(2):267-281. [doi: 10.1111/j.1939-0025.2010.01030.x] [Medline: 20553520]
17. Corrigan PW, Larson JE, Rüsch N. Self-stigma and the "why try" effect: impact on life goals and evidence-based practices. World Psychiatry 2009 Jun;8(2):75-81 [FREE Full text] [doi: 10.1002/j.2051-5545.2009.tb00218.x] [Medline: 19516923]
18. Pinfold V, Thornicroft G, Huxley P, Farmer P. Active ingredients in anti-stigma programmes in mental health. Int Rev Psychiatry 2005 Apr;17(2):123-131. [doi: 10.1080/09540260500073638] [Medline: 16194782]
19. Corrigan PW, Morris SB, Michaels PJ, Rafacz JD, Rüsch N. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. Psychiatr Serv 2012 Oct;63(10):963-973. [doi: 10.1176/appi.ps.201100529] [Medline: 23032675]
20. Clement S, van Nieuwenhuizen A, Kassam A, Flach C, Lazarus A, de Castro M, et al. Filmed v. live social contact interventions to reduce stigma: randomised controlled trial. Br J Psychiatry 2012 Jul;201(1):57-64. [doi: 10.1192/bjp.bp.111.093120] [Medline: 22157800]
21. Corrigan PW, Larson J, Sells M, Niessen N, Watson AC. Will filmed presentations of education and contact diminish mental illness stigma? Community Ment Health J 2007 Apr;43(2):171-181. [doi: 10.1007/s10597-006-9061-8] [Medline: 16988883]
22. Kalyanaraman SS, Penn DL, Ivory JD, Judge A. The virtual doppelganger: effects of a virtual reality simulator on perceptions of schizophrenia. J Nerv Ment Dis 2010 Jun;198(6):437-443. [doi: 10.1097/NMD.0b013e3181e07d66] [Medline: 20531123]
23. Kilteni K, Groten R, Slater M. The sense of embodiment in virtual reality. Presence Teleoperators Virtual Environments 2012 Nov;21(4):373-387. [doi: 10.1162/pres_a_00124]
24. Peck TC, Seinfeld S, Aglioti SM, Slater M. Putting yourself in the skin of a black avatar reduces implicit racial bias. Conscious Cogn 2013 Sep;22(3):779-787. [doi: 10.1016/j.concog.2013.04.016] [Medline: 23727712]

XSL•FO
RenderX

25. Petkova VI, Khoshnevis M, Ehrsson HH. The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. Front Psychol 2011;2:35. [doi: 10.3389/fpsyg.2011.00035] [Medline: 21687436]
26. Yee N, Bailenson JN. Walk a mile in digital shoes: the impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments. 2006 Presented at: Presence 2006: The 9th Annual International Workshop on Presence; August 24-26, 2006; Cleveland, OH.
27. Ando S, Clement S, Barley EA, Thornicroft G. The simulation of hallucinations to reduce the stigma of schizophrenia: a systematic review. Schizophr Res 2011 Dec;133(1-3):8-16. [doi: 10.1016/j.schres.2011.09.011] [Medline: 22005017]
28. Brown SA. Implementing a brief hallucination simulation as a mental illness stigma reduction strategy. Community Ment Health J 2010 Oct;46(5):500-504. [doi: 10.1007/s10597-009-9229-0] [Medline: 19669675]
29. Brown SA, Evans Y, Espenschade K, O'Connor M. An examination of two brief stigma reduction strategies: filmed personal contact and hallucination simulations. Community Ment Health J 2010 Oct;46(5):494-499. [doi: 10.1007/s10597-010-9309-1] [Medline: 20414723]
30. Cangas AJ, Navarro N, Parra JMA, Ojeda JJ, Cangas D, Piedra JA, et al. Stigma-Stop: a serious game against the stigma toward mental health in educational settings. Front Psychol 2017;8:1385. [doi: 10.3389/fpsyg.2017.01385] [Medline: 28878702]
31. Fisher S, Cleveland SE. Body Image and Personality. 2nd edition. Michigan: Dover Publications; 1968.
32. Biocca F. The cyborg's dilemma: progressive embodiment in virtual environments. J Computer-Mediated Commun 1997 Sep 01;3(2):1. [doi: 10.1111/j.1083-6101.1997.tb00070.x]
33. Maister L, Slater M, Sanchez-Vives MV, Tsakiris M. Changing bodies changes minds: owning another body affects social cognition. Trends Cogn Sci 2015 Jan;19(1):6-12. [doi: 10.1016/j.tics.2014.11.001] [Medline: 25524273]
34. Green MC, Kass S, Carrey J, Herzig B, Feeney R, Sabini J. Transportation across media: repeated exposure to print and film. Media Psychol 2008 Dec 23;11(4):512-539. [doi: 10.1080/15213260802492000]
35. Tal-Or N, Cohen J. Understanding audience involvement: conceptualizing and manipulating identification and transportation. Poetics 2010 Aug;38(4):402-418. [doi: 10.1016/j.poetic.2010.05.004]
36. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [doi: 10.1037//0022-3514.79.5.701] [Medline: 11079236]
37. Caputo NM, Rouner D. Narrative processing of entertainment media and mental illness stigma. Health Commun 2011 Oct;26(7):595-604. [doi: 10.1080/10410236.2011.560787] [Medline: 21516556]
38. Mann CE, Himelein MJ. Putting the person back into psychopathology: an intervention to reduce mental illness stigma in the classroom. Soc Psychiatry Psychiatr Epidemiol 2008 Jul;43(7):545-551. [doi: 10.1007/s00127-008-0324-2] [Medline: 18286216]
39. Oliver MB, Dillard JP, Bae K, Tamul DJ. The effect of narrative news format on empathy for stigmatized groups. Journalism Mass Commun Q 2012 Mar 14;89(2):205-224. [doi: 10.1177/1077699012439020]
40. Winterbottom A, Bekker HL, Conner M, Mooney A. Does narrative information bias individual's decision making? A systematic review. Soc Sci Med 2008 Dec;67(12):2079-2088. [doi: 10.1016/j.socscimed.2008.09.037] [Medline: 18951673]
41. Kaufman GF, Libby LK. Changing beliefs and behavior through experience-taking. J Pers Soc Psychol 2012 Jul;103(1):1-19. [doi: 10.1037/a0027525] [Medline: 22448888]
42. Appel M, Richter T. Transportation and need for affect in narrative persuasion: a mediated moderation model. Media Psychol 2010 May 28;13(2):101-135. [doi: 10.1080/15213261003799847]
43. Green MC, Brock TC, Kaufman GF. Understanding media enjoyment: the role of transportation into narrative worlds. Commun Theory 2004 Nov;14(4):311-327. [doi: 10.1111/j.1468-2885.2004.tb00317.x]
44. Schutte NS, Stilinović EJ. Facilitating empathy through virtual reality. Motivation Emotion 2017 Oct 16;41(6):708-712. [doi: 10.1007/s11031-017-9641-7]
45. Mak WWS, Chong ESK, Wong CCY. Beyond attributions: understanding public stigma of mental illness with the common sense model. Am J Orthopsychiatry 2014 Mar;84(2):173-181. [doi: 10.1037/h0099373] [Medline: 24826933]
46. Ehrsson HH. The experimental induction of out-of-body experiences. Science 2007 Aug 24;317(5841):1048 [FREE Full text] [doi: 10.1126/science.1142175] [Medline: 17717177]
47. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. PLoS One 2010 May 12;5(5):e10564 [FREE Full text] [doi: 10.1371/journal.pone.0010564] [Medline: 20485681]
48. Longo MR, Schüür F, Kammers MPM, Tsakiris M, Haggard P. What is embodiment? A psychometric approach. Cognition 2008 Jun;107(3):978-998. [doi: 10.1016/j.cognition.2007.12.004] [Medline: 18262508]
49. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: The Guilford Press; 2013.
50. Lysaker PH, Davis LW, Warman DM, Strasburger A, Beattie N. Stigma, social function and symptoms in schizophrenia and schizoaffective disorder: associations across 6 months. Psychiatry Res 2007 Jan 15;149(1-3):89-95. [doi: 10.1016/j.psychres.2006.03.007] [Medline: 17156853]
51. Sarge MA, Kim H, Velez JA. An intervention: the role of perspective taking in combating public stigma with virtual simulations. Cyberpsychol Behav Soc Netw 2020 Jan;23(1):41-51. [doi: 10.1089/cyber.2019.0678] [Medline: 31976769]

XSL•FO
RenderX

52. Rohrer MW. Seeing is believing: the importance of visualization in manufacturing simulation. 2000 Presented at: The 32nd Conference on Winter Simulation; December 2000; Orlando, FL. [doi: 10.1109/wsc.2000.899087]
53. Corrigan PW, Bink AB, Fokuo JK, Schmidt A. The public stigma of mental illness means a difference between you and me. Psychiatry Res 2015 Mar 30;226(1):186-191. [doi: 10.1016/j.psychres.2014.12.047] [Medline: 25660735]
54. Reinke RR, Corrigan PW, Leonhard C, Lundin RK, Kubiak MA. Examining two aspects of contact on the stigma of mental illness. J Soc Clin Psychol 2004 Jun;23(3):377-389. [doi: 10.1521/jscp.23.3.377.35457]
55. Thornicroft G, Rose D, Kassam A, Sartorius N. Stigma: ignorance, prejudice or discrimination? Br J Psychiatry 2007 Mar;190:192-193. [doi: 10.1192/bjp.bp.106.025791] [Medline: 17329736]
56. Gillath O, McCall C, Shaver PR, Blascovich J. What can virtual reality teach us about prosocial tendencies in real and virtual environments? Media Psychol 2008 Jun 25;11(2):259-282. [doi: 10.1080/15213260801906489]

Abbreviations

ANOVA: analysis of variance
IVR: immersive virtual reality

Edited by R Kukafka; submitted 26.08.20; peer-reviewed by C Stunden, A Algarin, S van Beugen; comments to author 06.10.20; revised version received 29.12.20; accepted 24.05.21; published 14.07.21.

Please cite as:
Yuen ASY, Mak WWS
The Effects of Immersive Virtual Reality in Reducing Public Stigma of Mental Illness in the University Population of Hong Kong: Randomized Controlled Trial
J Med Internet Res 2021;23(7):e23683
URL: https://www.jmir.org/2021/7/e23683
doi:10.2196/23683
PMID:34259636

©Anna S Y Yuen, Winnie W S Mak. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Resuscitating Cardiopulmonary Resuscitation Training in a Virtual Reality: Prospective Interventional Study

Janaya Elizabeth Perron[1], BMed, MD; Michael Jonathon Coffey[1], BMed, MD, PhD; Andrew Lovell-Simons[2], BA, MA; Luis Dominguez[3]; Mark E King[3], PhD; Chee Y Ooi[1], Dip Paeds, MBBS, PhD

[1]Discipline of Paediatrics, School of Women's and Children's Health, University of New South Wales, Randwick, Australia

[2]Medicine Education Development Unit, Faculty of Medicine, University of New South Wales, Sydney, Australia

[3]Educational Delivery Services, Office of the Pro Vice-Chancellor (Education), University of New South Wales, Sydney, Australia

Corresponding Author:
Janaya Elizabeth Perron, BMed, MD
Discipline of Paediatrics
School of Women's and Children's Health
University of New South Wales
Centre for Child Health Research & Innovation
Level 8, Bright Alliance Building, Corner of Avoca & High Streets
Randwick
Australia
Phone: 61 2 9382 5500
Fax: 61 2 9382 5681
Email: j.perron@unswalumni.com

Abstract

Background: Simulation-based technologies are emerging to enhance medical education in the digital era. However, there is limited data for the use of virtual reality simulation in pediatric medical education. We developed Virtual Doc as a highly immersive virtual reality simulation to teach pediatric cardiopulmonary resuscitation skills to medical students.

Objective: The primary objectives of this study were to evaluate participant satisfaction and perceived educational efficacy of Virtual Doc. The secondary aim of this study was to assess the game play features of Virtual Doc.

Methods: We conducted a prospective closed beta-testing study at the University of New South Wales (Sydney, Australia) in 2018. All medical students from the 6-year undergraduate program were eligible to participate and were recruited through voluntary convenience sampling. Participants attended a 1-hour testing session and attempted at least one full resuscitation case using the virtual reality simulator. Following this, participants were asked to complete an anonymous postsession questionnaire. Responses were analyzed using descriptive statistics.

Results: A total of 26 participants were recruited, consented to participate in this study, and attended a 1-hour in-person closed beta-testing session, and 88% (23/26) of participants completed the anonymous questionnaire and were included in this study. Regarding participant satisfaction, Virtual Doc was enjoyed by 91% (21/23) of participants, with 74% (17/23) intending to recommend the simulation to a colleague and 66% (15/23) intending to recommend the simulation to a friend. In assessment of the perceived educational value of Virtual Doc, 70% (16/23) of participants agreed they had an improved understanding of cardiopulmonary resuscitation, and 78% (18/23) agreed that Virtual Doc will help prepare for and deal with real-life clinical scenarios. Furthermore, 91% (21/23) of participants agreed with the development of additional Virtual Doc cases as beneficial for learning. An evaluation of the game play features as our secondary objective revealed that 70% (16/23) of participants agreed with ease in understanding how to use Virtual Doc, and 74% (17/23) found the game play elements useful in understanding cardiopulmonary resuscitation. One-third (7/23, 30%) found it easy to work with the interactive elements. In addition, 74% (17/23) were interested in interacting with other students within the simulation.

Conclusions: Our study demonstrates a positive response regarding trainee satisfaction and perceived educational efficacy of Virtual Doc. The simulation was widely accepted by the majority of users and may have the potential to improve educational learning objectives.

(J Med Internet Res 2021;23(7):e22920) doi:10.2196/22920

XSL•FO
RenderX

KEYWORDS

pediatrics; cardiopulmonary resuscitation; virtual reality; medical education

Introduction

Simulation-based education plays a significant role in developing vital clinical assessment and management skills, especially in emergency scenarios such as a sudden cardiac arrest (SCA). SCA is an uncommon but life-threatening condition within the pediatric population [1,2] that includes the potential catastrophic consequences of mortality and poor neurological morbidity. An estimated 16,000 [3,4] children in the United States experience out-of-hospital SCAs and 5800 [4] experience in-hospital SCAs every year. Overall, survival rates are better for in-hospital arrests compared with out-of-hospital arrests [5], which is unsurprising considering that the most significant predictor of survival is early initiation of adequate cardiopulmonary resuscitation (CPR) [6]. However, junior doctors often have limited real-world learning experience managing an acutely ill and vulnerable child [7], especially in emergent scenarios, where active participation is often reserved for more experienced practitioners. In addition, many pediatric health care professionals feel inadequately prepared to manage a critically ill child, increasing the risk of medical errors [8]. Therefore, simulation training forms the cornerstone of resuscitation education, providing an opportunity for trainees to prepare for real-world clinical practice without risking patient safety [9-11]. However, in-person simulation training sessions are expensive and resource-intensive [12], and, therefore, the efficacy of alternative modalities should be explored.

The revolutionary development of highly immersive digitized learning resources such as virtual reality simulators harness the advantages of simulation while offering a cost-effective and widely accessible educational platform. Several studies have provided evidence for the use of virtual reality simulators for procedural skills training in the surgical field, with improved intraoperative performance [13] and accuracy [14] and a reduction in operating times [13] and errors [15-18]. One study investigated the use of virtual reality for CPR education compared with traditional mannequins and found virtual reality could be a valid and acceptable training method [19]. Furthermore, in a study by Wong et al [20], clinical CPR instructors outlined the limits of traditional education and noted the potential beneficial features of virtual reality as an additional learning tool. The experiential nature of virtual reality may also be conducive to developing procedural skills such as CPR, as repetitive hands-on practice permits the development of muscle memory [21], leading to technical skill competence. CPR involves the activation of declarative memory to recall the sequential steps and procedural memory to perform the active steps such as chest compressions [22], both of which may be developed and reinforced through the use of virtual reality.

To the best of our knowledge, there are no highly immersive virtual reality simulators for pediatric CPR training. As such, we developed Virtual Doc, a highly immersive and experiential 3D multimedia sensory virtual environment designed to teach pediatric CPR skills to medical students and junior doctors. We performed a prospective closed beta-testing study on undergraduate medical students with the primary aim to evaluate participant satisfaction and perceived educational efficacy of Virtual Doc. The secondary aim of this study was to assess the gameplay features of our simulation. We hypothesized that Virtual Doc would be an enjoyable and highly educational learning experience. We also hypothesized that the gameplay features would be beneficial in the facilitation of a realistic hands-on clinical learning experience.

Methods

Study Design

We conducted a prospective, uncontrolled, interventional closed beta-testing study in 2018 to assess the usability, acceptability, and perceived educational effectiveness of Virtual Doc as a method for teaching pediatric CPR to medical students. All participants provided implied consent by expressing interest in the study, attending a 1-hour closed beta-testing session, and completing an anonymous postsession questionnaire (Multimedia Appendix 1). The study was approved by the University of New South Wales Human Research Ethics Committee (HC180484).

Participants

Participants were eligible to partake in this study if they were currently enrolled in the 6-year undergraduate medical program at the University of New South Wales, Sydney, Australia. All medical students from the 6-year program were eligible and invited to participate in this study through an information email sent out by the University of New South Wales Medical Society administration. Students were instructed to contact the study investigators with their expression of interest. They were subsequently allocated to a 1-hour session in which they would engage with Virtual Doc and complete the postsession questionnaire. Participants were excluded from the analysis if they did not complete the questionnaire.

Virtual Doc

Virtual Doc is a virtual reality simulation developed for Oculus VR (Facebook Technologies LLC; Figure 1). The Virtual Doc app was developed using Unity3D software (Unity Technologies), and models were developed using open-source Blender modeling software. The simulation was conducted using the hardware of the Oculus Rift system linked to Alienware laptops (Alienware Corp). Virtual Doc is a first-person active learning experience through immersion within a multimedia sensory environment. Users wear a headset and use 2 hand controllers and are able to use these controllers, which are equipped with haptic technologies, to interact with the virtual environment. This simulation enables users to perform a series of actions including picking up objects, pushing buttons, and turning dials. They are also fully immersed with audiovisual aids such as auscultation of the heart and lungs.

XSL•FO
RenderX

Figure 1. The Virtual Doc environment.



The scenario in this study involved an SCA in a young child that required appropriate and timely management. The required cardiac arrest management was in keeping with the Advanced Pediatric Life Support Australia guidelines [23]. Students were provided with a brief written clinical history on the main screen in the virtual environment and were then instructed to begin CPR. To complete the case, participants were required to assess the surroundings, check for a response, signal for senior help by pressing a call button, and then begin resuscitation by assessing the airway, breathing, and circulation; bagging and masking the patient (Figure 2); performing adequate chest compressions; and defibrillating the patient. As this scenario featured a shockable rhythm, this patient was defibrillated as part of the SCA management algorithm; however, this step may not be required for all pediatric SCA presentations. Shortly after the participant signaled for help, a senior physician arrived in the virtual environment to resume chest compressions while the user prepared for defibrillation. For successful completion of the case, all steps must have been completed correctly within 10 minutes.

Figure 2. User is ventilating the patient with a bag-valve mask as per the resuscitation algorithm.



Outcome Measures

Validating the use of highly immersive and experiential virtual reality simulators as a form of active learning can be achieved by assessing various outcomes through the Kirkpatrick model of evaluation [24]. This 4-tiered hierarchy includes a scaffolding evaluation of reaction (level 1), learning (level 2), behavior (level 3), and results (level 4) [24]. As the levels progress, there

XSL•FO
RenderX

is an increase in conclusively defining the efficacy of the educational activity [25].

The primary outcomes of this study were participant satisfaction (Kirkpatrick level 1) and the perceived educational value of Virtual Doc (Kirkpatrick level 2). Participant satisfaction was assessed through 1 Likert-style question and 2 trichotomous questions regarding enjoyment of the game and whether the participant would recommend the game to a colleague or friend. The perceived educational efficacy was assessed through 3 Likert-style questions including whether participants improved their understanding of the simulation objectives, whether engaging with Virtual Doc prepares students for a similar real-life clinical scenario, and whether different clinical cases would be beneficial for learning.

The secondary outcome was an evaluation of the gameplay features. Here, subjects were asked to respond to 3 Likert-style questions and 1 dichotomous question regarding the ease of understanding how to engage with the simulation, the level of ease to work with the interactive elements, whether the interactive elements were useful for learning, and whether the user would like to interact with other students in the virtual world.

Testing Session

All interested participants were allocated to a 1-hour in-person session. Attendees were invited to attempt at least one full clinical scenario using the virtual reality simulator. They were offered an unlimited number of attempts within the 1-hour session. After engaging with Virtual Doc, participants were invited to complete an anonymous questionnaire that evaluated the primary and secondary outcomes.

Data Collection and Statistical Analysis

Each postsession questionnaire was assigned a unique identification number and was completed anonymously by participants. The responses were collated and outcome measures were analyzed using descriptive statistics. The 7-point Likert scale responses were streamlined to 3 categories of agree (included all responses indicating strongly agree, agree, and mildly agree), neutral, and disagree (included all responses indicating mildly disagree, disagree, and strongly disagree) for the data analysis. Trichotomous questions requiring a response of yes, not sure, or no and dichotomous questions requiring a response of yes or no were analyzed as such. The data were analyzed using the statistical software R (version 1.1.423, R Foundation for Statistical Computing) [26].

Results

A total of 26 participants were recruited and attended 1-hour in-person closed beta-testing sessions. A total of 88% (23/26) of participants completed the anonymous questionnaire and were included in the analysis, and 12% (3/26) attended the testing session but declined to complete the questionnaire and were therefore excluded from the analysis. They were not required to provide a reason for declining to participate.

Study Population

All 23 of the included participants completed the demographic characteristics section of the questionnaire. Overall, the median age of participants was 22.0 (interquartile range 21.5-23.0) years, and 70% (16/23) of participants were female. In 2018, 61% (14/23) of participants were in year 4, 13% (3/23) were in year 5, and 26% (6/23) were in year 6. A total of 87% (20/23) of participants were local students, while 13% (3/23) were international students.

Outcome Measures

Participants responded to 10 Likert-style, trichotomous, or dichotomous questions regarding their experience with Virtual Doc. The results of these questions are summarized by question type in the horizontal stacked bar graphs in Figures 3-5.

XSL•FO
RenderX

Figure 3. Results of the Likert-style questions.

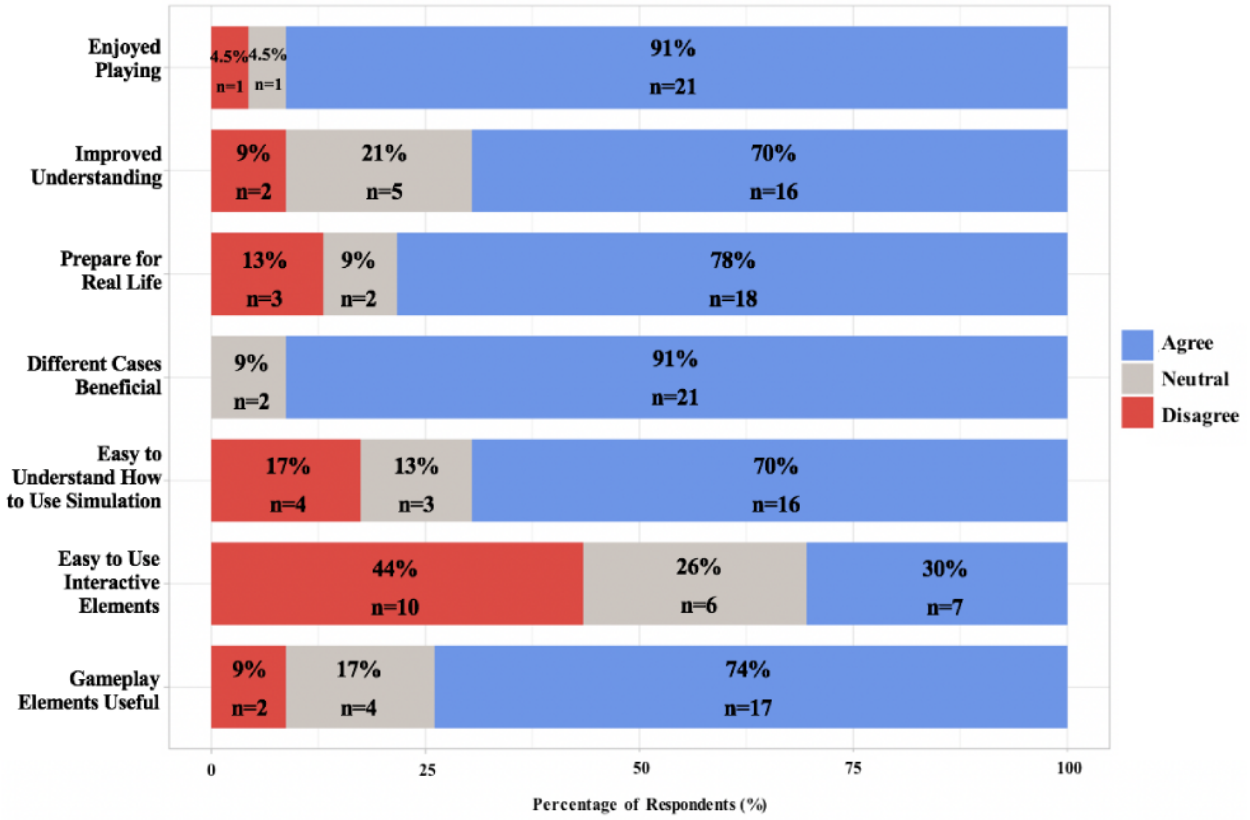


Figure 4. Results of the trichotomous questions.

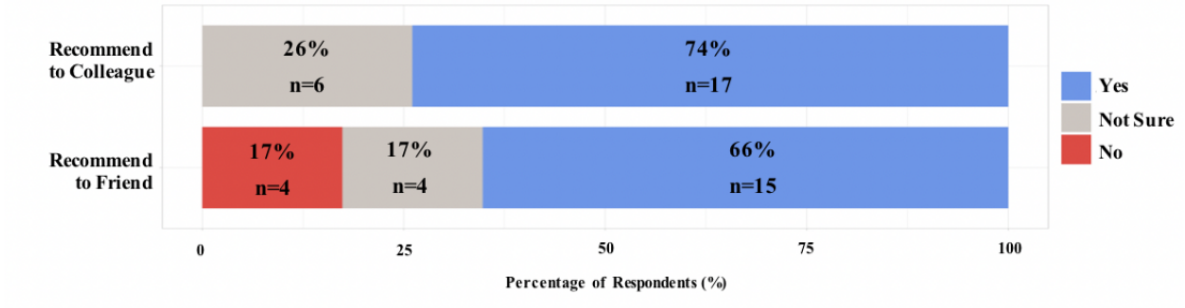
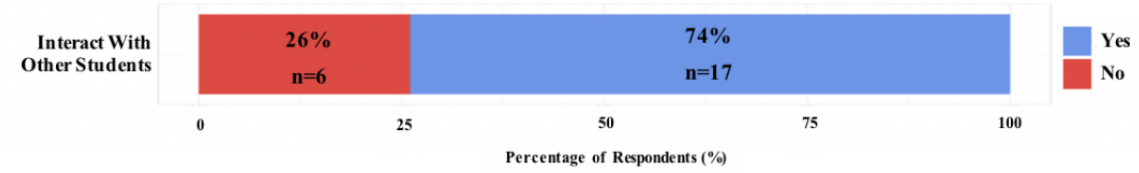


Figure 5. Results of the dichotomous question.



Participant Satisfaction

Respondents completed 1 Likert-style question and 2 trichotomous questions to evaluate satisfaction with Virtual Doc as an assessment of Kirkpatrick level 1. Virtual Doc was enjoyed by 91% (21/23) of participants. Furthermore, 74% (17/23) would recommend this simulation to a colleague, and 66% (15/23) would recommend this simulation to a friend.

Perceived Educational Value

Participants completed 3 Likert-style questions to assess the perceived educational efficacy of Virtual Doc, evaluating Kirkpatrick level 2, and 70% (16/23) of participants agreed they had an improved understanding of CPR following the use of Virtual Doc. In addition, 78% (18/23) agreed that Virtual Doc will help them prepare for and deal with real-life clinical scenarios. Furthermore, 91% (21/23) agreed there was a learning benefit to developing different cases within Virtual Doc, and 9% (2/23) were neutral.

XSL•FO
RenderX

Gameplay Features

Respondents completed 3 Likert-style questions and 1 dichotomous question evaluating the gameplay features of Virtual Doc. In terms of usability and gameplay, 70% (16/23) of participants agreed they had ease in understanding how to use Virtual Doc, and 74% (17/23) found the gameplay elements useful for understanding CPR. Furthermore, 30% (7/23) of participants found it easy to work with the interactive elements, which provides a template for future software improvements. In addition, 74% (17/23) would like the option to interact with other students within the game.

Discussion

Principal Findings

To the best of our knowledge, a prospective study on the use of a virtual reality simulator to teach pediatric CPR to medical students has yet to be conducted. In accordance with the Kirkpatrick model [24], the results of our study demonstrated high trainee satisfaction (Kirkpatrick level 1) and a perceived improvement in learning (Kirkpatrick level 2). In evaluating our secondary outcome of assessing the response to the virtual reality gameplay features, the majority of participants agreed with the ease in understanding how to use the simulation and the usefulness of gameplay features in improving their understanding of CPR and expressed a desire for interacting with others in the virtual environment. Although some participants found it difficult to work with the interactive elements, this closed beta test provided an opportunity to identify elements for future improvements, including the refinement of the interactive elements and future considerations such as a multiplayer design which may further enhance the learning experience and element of realism given the significant role of teamwork in resuscitation. Current data on trainee satisfaction, educational efficacy, and gameplay elements of virtual reality simulators for pediatric medical education is limited, and the encouraging results of this study will address the gap in the literature and could have an integral role in transforming the culture of future medical education.

Immersive technologies provide users with an engaging and enjoyable learning platform to train skills to proficiency without risking patient safety. Our study illustrated high participant satisfaction as Virtual Doc was enjoyed by almost all participants. This result parallels a study by Kron et al [27] in which an evaluation of medical student attitudes toward technology-based education was conducted and revealed that almost all students liked the idea of using technology to enhance their educational experience and thought that education should make better use of new media technologies. Beyond satisfaction, a previous study reported that experiential simulations offer a high level of interactivity and engagement, which may increase user motivation to learn [28]. Providing a highly engaging environment, such as that of Virtual Doc, is important for trainees as increased interest and satisfaction with the learning modality may result in an increased level of motivation to learn; one study showed that highly motivated students are more effective learners [29]. The importance of satisfaction and effective learning is reinforced by a study evaluating the effectiveness of virtual reality versus traditional or other forms of digital learning in educating health care professionals, in which a higher level of interactivity was more effective for the development of postintervention knowledge and skills [30]. The level of satisfaction with a learning experience also affects the probability of a user recommending such an experience to their peers [31]. The results of our study are congruent with this notion, as indicated by most students enjoying the simulation and agreeing with the intention of recommendation.

In this study, Virtual Doc was perceived as educationally efficacious by the majority of the participants, as indicated by a positive response toward improved understanding, preparation for real-world clinical scenarios, and support for designing additional cases. An important component of effective education in the adult learner involves the psychological concept of flow. This term is defined as a state in which a learner is “fully engaged, focused, and committed to the success of the activity” [32], and this can be achieved through active engagement in an experiential learning process. The use of highly immersive and interactive virtual reality simulators can improve flow, which may lead to superior learning outcomes and performance as echoed by the majority of participants reporting an improved understanding of CPR. Additionally, engaging sensory modalities including vision, haptics, and audition also promotes active learning, which can improve memory retention [33]. Virtual Doc includes sensory features through auscultation of heart and lung sounds and palpation of the peripheral pulses, which the user will hopefully be able to translate into the real-world clinical setting when assessing or managing a similar patient presentation.

Being immersed in the virtual yet realistic hospital environment of Virtual Doc provides users with a lifelike clinical experience. This digital-based environment is an innovative tool with realistic features supporting the recommendations of adult learning theories such as Knowles’ theory [34,35]. The responses in this study regarding the Virtual Doc environment and its ability to prepare students for real-life clinical scenarios showcases a facility in which the elements of adult learning theories can be included. Active experimentation within a realistic and immersive learning environment allows users to make clinical decisions and experience the consequences of their actions in real time, mimicking a real-world clinical environment. This element of realism is important in encouraging intrinsic motivation to support the needs of an adult learner.

Additionally, the Kolb cycle of learning encourages repetition, reflection, and correction to improve learning outcomes [36]. Virtual Doc supports this cyclical relationship as learners have the opportunity to engage in repetitive practice until they achieve educational mastery in a safe and controlled environment, which is an important benefit of simulation as this is severely limited in real-world emergency patient encounters, especially in the pediatric population [5]. The flexible design of virtual reality may enable health care professionals to have more convenient and frequent opportunities to practice and refine skills.

Several studies have demonstrated that junior doctors feel inadequately prepared and lack confidence in their early training

XSL•FO
RenderX

years [37-41]. This is further supplemented by less favorable outcomes for neonatal resuscitation being described in the summer when senior staff physicians are less likely to be present and the overall volume of staff members is reduced [42]. Lacking confidence and inadequate preparation may translate into unsafe practice and poor patient outcomes and is an especially important consideration when managing emergency scenarios such as an SCA. Virtual reality simulators such as Virtual Doc may provide an avenue to bridge the gap between trainee and doctor through immersive gameplay by enabling users to master foundational principles and basic management algorithms and apply this knowledge to challenging simulated cases. In our study, the majority of respondents expressed that this simulation game would aid in developing clinical skills and preparing the user for dealing with real-life clinical scenarios. In addition, almost all participants expressed benefit in the development of different cases, indicating a positive response toward the clinical educational value of the simulation.

The gameplay elements involved in virtual reality add entertainment to the learning process, thereby motivating learners to engage in study. The triad of immersion, interaction, and imagination, as described by Burdea and Coiffet [43], are important factors of virtual reality technology. Coupled with andragogical principles, a highly interactive game design improves intrinsic motivation and creates a more enjoyable and engaging learning experience with a consequential potential to illustrate superior examination scores [44]. In our study, the vast majority of respondents agreed that the interactive gameplay elements were useful in understanding how to perform CPR and found the process of the virtual reality simulation easy to understand. This was further supplemented by almost all participants enjoying the experience in the assessment of participant satisfaction. However, some respondents did not experience ease in using the interactive elements within the virtual environment. As this was a closed beta test, the software was able to deliver its intentional value to the participants, but there were some minor issues reported with ease in fully integrating into the environment such as accurately picking up a stethoscope or pushing the assistance button. This feedback is being considered as we progress toward the final product and acts as identified areas for future developments.

Finally, nontechnical skills such as teamwork and leadership are paramount in successful resuscitation. The collegiality of the responding team directly impacts predictors of postarrest survival as a coordinated course of action is associated with improved patient outcomes [45]. Technology-based games such as Virtual Doc have the potential to facilitate an atmosphere of teamwork, and in the digital era, medical students have agreed with introducing multiplayer simulations if they were fun and developed clinical skills [27]. Likewise, our study demonstrated that the majority of participants were interested in interacting with other players. A multiplayer design within the virtual environment could afford an opportunity to rotate through different resuscitation team roles, supporting an active and immersive learning experience, with the potential to equip learners with the crucial teamwork skills required for the effective management of a pediatric SCA.

Limitations

Our study was limited by the small sample size and monocentric design, which limits its generalizability and external validity. However, participants from all years of study were eligible to participate, providing a degree of generalizability. This trial is also limited by voluntary convenience sampling, leading to a potential selection bias. Furthermore, this study used a single intervention with no active comparator, and therefore, the participants were not randomized and were unblinded to the intervention, which may formulate a measurement bias. However, the questionnaires were anonymized, providing a layer of security for participants to provide uncoerced feedback. Moreover, the final version of Virtual Doc will differ from the prototype used in this study as this was closed beta-testing and we are improving the software to address the feedback provided by the participants in this study. In addition, the prototype design of Virtual Doc in this closed beta test did not assess user hand technique or the quality of chest compressions. This limitation will guide future software improvements to ensure users are equipped with real-time visual and haptic feedback regarding the technique and quality of chest compressions.

Future Research

Given the encouraging results of this study, future research should include an investigator-blinded randomized controlled trial to objectively evaluate the educational efficacy of Virtual Doc against traditional simulation-based education or other digitized modalities such as serious games or e-learning modules. Furthermore, an evaluation of translation of knowledge into clinical practice (Kirkpatrick level 3) and the impact of learned knowledge on changes within the organizational practice to improve patient outcomes (Kirkpatrick level 4) should be performed, with emphasis on a reduction in medical errors and improved patient survival rates and overall health outcomes. This study does not assess the higher levels of the Kirkpatrick model, and this limitation should be the focus of future research to contribute to the quality of data for the efficacy of Virtual Doc.

Conclusion

In summary, our study demonstrates a positive response regarding the satisfaction and educational efficacy of Virtual Doc. Our findings reveal that this virtual reality simulation was widely accepted by the majority of users and has the potential to improve educational learning objectives. As such, our results provide a promising contribution to the educational revolution and may encourage the use of this emerging and versatile technology in the transformation of the 21st century medical curricula.

XSL•FO
RenderX

Acknowledgments

The study team would like to acknowledge the participating medical students for their involvement in this study, Reece Pahn (University of New South Wales Medical Society secretary) for facilitating the invitation of students, and Seiya Takeda and Zheyu Li for assisting with game development.

Authors' Contributions

The study was conceived by CYO and MJC. The study design and analysis were developed by CYO and MJC. CYO, MJC, ALS, and LD contributed to the conception and design of Virtual Doc. Subject recruitment and trial coordination were performed by JEP and MJC. LD sourced and coordinated the required technological equipment for the testing sessions. Data collection and statistical analysis were performed by JEP, MJC, and CYO. The paper was first drafted by JEP and revised in subsequent drafts by MJC, MEK, and CYO. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire instrument.
[DOC File , 236 KB - jmir_v23i7e22920_app1.doc]

References

1. Ilina MV, Kepron CA, Taylor GP, Perrin DG, Kantor PF, Somers GR. Undiagnosed heart disease leading to sudden unexpected death in childhood: a retrospective study. Pediatrics 2011 Sep;128(3):e513-e520. [doi: 10.1542/peds.2010-2307] [Medline: 21824887]
2. Meyer L, Stubbs B, Fahrenbruch C, Maeda C, Harmon K, Eisenberg M, et al. Incidence, causes, and survival trends from cardiovascular-related sudden cardiac arrest in children and young adults 0 to 35 years of age: a 30-year review. Circulation 2012 Sep 11;126(11):1363-1372. [doi: 10.1161/CIRCULATIONAHA.111.076810] [Medline: 22887927]
3. Topjian AA, Nadkarni VM, Berg RA. Cardiopulmonary resuscitation in children. Curr Opin Crit Care 2009 Jun;15(3):203-208. [doi: 10.1097/mcc.0b013e32832931e1] [Medline: 19469022]
4. Marino BS, Tabbutt S, MacLaren G, Hazinski MF, Adatia I, Atkins DL, American Heart Association Congenital Cardiac Defects Committee of the Council on Cardiovascular Disease in the Young, Council on Clinical Cardiology, Council on Cardiovascular and Stroke Nursing, Council on Cardiovascular Surgery and Anesthesia, Emergency Cardiovascular Care Committee. Cardiopulmonary resuscitation in infants and children with cardiac disease: a scientific statement from the American Heart Association. Circulation 2018 May 29;137(22):e691-e782. [doi: 10.1161/CIR.0000000000000524] [Medline: 29685887]
5. Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. JAMA 2006 Jan 4;295(1):50-57. [doi: 10.1001/jama.295.1.50] [Medline: 16391216]
6. Goto Y, Maeda T, Goto Y. Impact of dispatcher-assisted bystander cardiopulmonary resuscitation on neurological outcomes in children with out-of-hospital cardiac arrests: a prospective, nationwide, population-based cohort study. J Am Heart Assoc 2014 Apr 30;3(3):e000499 [FREE Full text] [doi: 10.1161/JAHA.113.000499] [Medline: 24785780]
7. Clerihew L, Rowney D, Ker J. Simulation in paediatric training. Arch Dis Child Educ Pract Ed 2016 Feb;101(1):8-14 [FREE Full text] [doi: 10.1136/archdischild-2015-309143] [Medline: 26614805]
8. Cheng A, Duff J, Grant E, Kissoon N, Grant VJ. Simulation in paediatrics: an educational revolution. Paediatr Child Health 2007 Jul;12(6):465-468 [FREE Full text] [doi: 10.1093/pch/12.6.465] [Medline: 19030409]
9. Ten Eyck RP. Simulation in emergency medicine training. Pediatr Emerg Care 2011 Apr;27(4):333-341. [doi: 10.1097/PEC.0b013e3182131fe0] [Medline: 21467889]
10. Issenberg SB, McGaghie WC, Petrusa ER, Lee GD, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005 Jan;27(1):10-28. [doi: 10.1080/01421590500046924] [Medline: 16147767]
11. Gaba DM. The future vision of simulation in health care. Qual Saf Health Care 2004 Oct;13 Suppl 1:i2-i10 [FREE Full text] [doi: 10.1136/qhc.13.suppl_1.i2] [Medline: 15465951]
12. Drummond D, Arnaud C, Thouvenin G, Guedj R, Grimprel E, Duguet A, et al. An innovative pedagogic course combining video and simulation to teach medical students about pediatric cardiopulmonary arrest: a prospective controlled study. Eur J Pediatr 2016 Jun;175(6):767-774. [doi: 10.1007/s00431-016-2702-1] [Medline: 26847430]
13. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 2009 May 14;338:b1802 [FREE Full text] [doi: 10.1136/bmj.b1802] [Medline: 19443914]

XSL•FO
RenderX

14. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 2005 Feb;241(2):364-372. [doi: 10.1097/01.sla.0000151982.85062.80] [Medline: 15650649]
15. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: is one superior to the other? Surg Endosc 2004 Mar;18(3):485-494. [doi: 10.1007/s00464-003-9043-7] [Medline: 14752633]
16. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004 Feb;91(2):146-150. [doi: 10.1002/bjs.4407] [Medline: 14760660]
17. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-463. [doi: 10.1097/01.SLA.0000028969.51489.B4] [Medline: 12368674]
18. McClusky DA, Gallagher AG, Ritter E, Lederman AB, Van Sickle KR, Baghai M, et al. Virtual reality training improves junior residents' operating room performance: results of a prospective, randomized, double-blinded study of the complete laparoscopic cholecystectomy. J Am Coll Surg 2004 Sep;199(3):73. [doi: 10.1016/j.jamcollsurg.2004.05.157]
19. Semeraro F, Ristagno G, Giulini G, Gnudi T, Kayal JS, Monesi A, et al. Virtual reality cardiopulmonary resuscitation (CPR): comparison with a standard CPR training mannequin. Resuscitation 2019 Feb;135:234-235. [doi: 10.1016/j.resuscitation.2018.12.016] [Medline: 30597133]
20. Wong M, Chue S, Jong M, Benny H, Zary N. Clinical instructors' perceptions of virtual reality in health professionals' cardiopulmonary resuscitation education. SAGE Open Med 2018;6:2050312118799602 [FREE Full text] [doi: 10.1177/2050312118799602] [Medline: 30245815]
21. Hunt EA, Duval-Arnould JM, Nelson-McMillan KL, Bradshaw JH, Diener-West M, Perretta JS, et al. Pediatric resident resuscitation skills improve after "rapid cycle deliberate practice" training. Resuscitation 2014 Jul;85(7):945-951. [doi: 10.1016/j.resuscitation.2014.02.025] [Medline: 24607871]
22. Sense F, Maass S, van Rijn H. Interactions of declarative and procedural memory in real-life tasks: validating CPR as a new paradigm. 2016 Presented at: Proceedings of the 14th International Conference on Cognitive Modeling; 2016; University Park.
23. Advanced Life Support Group. Advanced Paediatric Life Support, Australia and New Zealand: A Practical Approach to Emergencies. 6th edition. West Sussex: John Wiley & Sons; 2017.
24. Kirkpatrick DL, Kirkpatrick J. Evaluating Training Programs: The Four Levels. 3rd Edition. San Francisco: Berrett-Koehler Publishers, Inc; 2006.
25. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul (Lond) 2017;2:3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
26. Hornik K. The R FAQ. URL: https://cran.r-project.org/doc/FAQ/R-FAQ.html [accessed 2019-09-13]
27. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ 2010;10:50 [FREE Full text] [doi: 10.1186/1472-6920-10-50] [Medline: 20576125]
28. Limniou M, Roberts D, Papadopoulos N. Full immersive virtual environment CAVETM in chemistry education. Comput Educ 2008 Sep;51(2):584-593. [doi: 10.1016/j.compedu.2007.06.014]
29. Sutcliffe A. Multimedia and Virtual Reality: Designing Multisensory User Interfaces. Mahwah: Psychology Press; 2003.
30. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
31. Hyejin C, Woonhan K, Hwayeol C. Effects of smartphone-based learning properties on user satisfaction and recommendation intention. Indian J Sci Technol 2015;8:26. [doi: 10.17485/ijst/2015/v8i26/80990]
32. Taekman JM, Shelley K. Virtual environments in healthcare: immersion, disruption, and flow. Int Anesthesiol Clin 2010;48(3):101-121. [doi: 10.1097/AIA.0b013e3181eace73] [Medline: 20616640]
33. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 2014 Jun 10;111(23):8410-8415 [FREE Full text] [doi: 10.1073/pnas.1319030111] [Medline: 24821756]
34. Akl EA, Sackett K, Pretorius R, Erdley S, Bhoopathi PS, Mustafa R, et al. Educational games for health professionals. Cochrane Database Syst Rev 2008 Jan 23(1):CD006411. [doi: 10.1002/14651858.CD006411.pub2] [Medline: 18254103]
35. Knowles M. The Modern Practice of Adult Education: Andragogy Verus Pedagogy. New York: Association Press; 1970.
36. Kolb AY, Kolb DA. Learning styles and learning spaces: enhancing experiential learning in higher education. Acad Manag Learn Educ 2005 Jun;4(2):193-212. [doi: 10.5465/amle.2005.17268566]
37. Tallentire VR, Smith SE, Skinner J, Cameron HS. The preparedness of UK graduates in acute care: a systematic literature review. Postgrad Med J 2012 Jul;88(1041):365-371. [doi: 10.1136/postgradmedj-2011-130232] [Medline: 22167809]
38. Goldacre MJ, Taylor K, Lambert TW. Views of junior doctors about whether their medical school prepared them well for work: questionnaire surveys. BMC Med Educ 2010 Nov 11;10:78 [FREE Full text] [doi: 10.1186/1472-6920-10-78] [Medline: 21070622]

XSL•FO
RenderX

39. Bligh J. The first year of doctoring: still a survival exercise. Med Educ 2002 Jan;36(1):2-3. [doi: 10.1046/j.1365-2923.2002.01129.x] [Medline: 11849515]
40. Wall D, Bolshaw A, Carolan J. From undergraduate medical education to pre-registration house officer year: how prepared are students? Med Teach 2006 Aug;28(5):435-439. [doi: 10.1080/01421590600625171] [Medline: 16973456]
41. Drummond D, Arnaud C, Thouvenin G, Guedj R, Duguet A, de Suremain N, et al. [Newly formed French residents in pediatrics are not well prepared for conducting pediatric resuscitation after medical school]. Arch Pediatr 2016 Feb;23(2):150-158. [doi: 10.1016/j.arcped.2015.11.012] [Medline: 26774839]
42. Stewart JH, Andrews J, Cartlidge PH. Numbers of deaths related to intrapartum asphyxia and timing of birth in all Wales perinatal survey, 1993-1995. BMJ 1998 Feb 28;316(7132):657-660 [FREE Full text] [doi: 10.1136/bmj.316.7132.657] [Medline: 9522787]
43. Burdea GC, Coiffet P. Virtual Reality Technology. 2nd Edition. Hoboken: John Wiley & Sons; 2003.
44. Kiili K. Digital game-based learning: towards an experiential gaming model. Internet Higher Educ 2005 Jan;8(1):13-24. [doi: 10.1016/j.iheduc.2004.12.001]
45. Hunziker S, Johansson AC, Tschan F, Semmer NK, Rock L, Howell MD, et al. Teamwork and leadership in cardiopulmonary resuscitation. J Am Coll Cardiol 2011 Jun 14;57(24):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2011.03.017] [Medline: 21658557]

Abbreviations

CPR: cardiopulmonary resuscitation
SCA: sudden cardiac arrest

Edited by G Eysenbach, R Kukafka; submitted 29.07.20; peer-reviewed by T Canares, R Lundin, D Drummond; comments to author 18.09.20; revised version received 16.10.20; accepted 11.03.21; published 29.07.21.

Please cite as:
Perron JE, Coffey MJ, Lovell-Simons A, Dominguez L, King ME, Ooi CY
Resuscitating Cardiopulmonary Resuscitation Training in a Virtual Reality: Prospective Interventional Study
J Med Internet Res 2021;23(7):e22920
URL: https://www.jmir.org/2021/7/e22920
doi:10.2196/22920
PMID:34326040

©Janaya Elizabeth Perron, Michael Jonathon Coffey, Andrew Lovell-Simons, Luis Dominguez, Mark E King, Chee Y Ooi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 29.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Recommendations for the Design and Implementation of Virtual Reality for Acquired Brain Injury Rehabilitation: Systematic Review

Sophie Brassel[1], BAppSc; Emma Power[2], PhD; Andrew Campbell[3], PhD; Melissa Brunner[1], PhD; Leanne Togher[1], PhD

[1]Discipline of Speech Pathology, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Speech Pathology, Graduate School of Health, University of Technology Sydney, Sydney, Australia

[3]Cyberpsychology Research Group, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Sophie Brassel, BAppSc
Discipline of Speech Pathology, Sydney School of Health Sciences
Faculty of Medicine and Health
The University of Sydney
Susan Wakil Health Building D18
Western Avenue, The University of Sydney
Sydney, 2006
Australia
Phone: 61 2 8627 9541
Email: sophie.brassel@sydney.edu.au

Abstract

Background: Virtual reality (VR) is increasingly being used for the assessment and treatment of impairments arising from acquired brain injuries (ABIs) due to perceived benefits over traditional methods. However, no tailored options exist for the design and implementation of VR for ABI rehabilitation and, more specifically, traumatic brain injury (TBI) rehabilitation. In addition, the evidence base lacks systematic reviews of immersive VR use for TBI rehabilitation. Recommendations for this population are important because of the many complex and diverse impairments that individuals can experience.

Objective: This study aims to conduct a two-part systematic review to identify and synthesize existing recommendations for designing and implementing therapeutic VR for ABI rehabilitation, including TBI, and to identify current evidence for using immersive VR for TBI assessment and treatment and to map the degree to which this literature includes recommendations for VR design and implementation.

Methods: This review was guided by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). A comprehensive search of 11 databases and gray literature was conducted in August 2019 and repeated in June 2020. Studies were included if they met relevant search terms, were peer-reviewed, were written in English, and were published between 2009 and 2020. Studies were reviewed to determine the level of evidence and methodological quality. For the first part, qualitative data were synthesized and categorized via meta-synthesis. For the second part, findings were analyzed and synthesized descriptively owing to the heterogeneity of data extracted from the included studies.

Results: In the first part, a total of 14 papers met the inclusion criteria. Recommendations for VR design and implementation were not specific to TBI but rather to stroke or ABI rehabilitation more broadly. The synthesis and analysis of data resulted in three key phases and nine categories of recommendations for designing and implementing VR for ABI rehabilitation. In the second part, 5 studies met the inclusion criteria. A total of 2 studies reported on VR for assessment and three for treatment. Studies were varied in terms of therapeutic targets, VR tasks, and outcome measures. VR was used to assess or treat impairments in cognition, balance, and anxiety, with positive outcomes. However, the levels of evidence, methodological quality, and inclusion of recommendations for VR design and implementation were poor.

Conclusions: There is limited research on the use of immersive VR for TBI rehabilitation. Few studies have been conducted, and there is limited inclusion of recommendations for therapeutic VR design and implementation. Future research in ABI rehabilitation should consider a stepwise approach to VR development, from early co-design studies with end users to larger controlled trials. A list of recommendations is offered to provide guidance and a more consistent model to advance clinical research in this area.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e26344) doi:10.2196/26344

KEYWORDS

virtual reality; acquired brain injury; traumatic brain injury; rehabilitation; systematic review; recommendations; cognitive communication; mobile phone

Introduction

Background

The use of virtual reality (VR) in health care has expanded in recent years and continues to be investigated due to the increasing availability and advancement of technology [1,2]. VR is being used in clinical research for the assessment and therapeutic intervention of impairments associated with acquired brain injuries (ABIs), which is an umbrella term for brain injuries that are sustained after birth [3,4]. A considerable evidence base exists for using VR for ABI rehabilitation, with a particular focus on stroke [5-14]. Traumatic brain injury (TBI) is another ABI that is increasingly being investigated with VR technologies; however, the evidence base is smaller and not as rigorous as that for other ABIs. TBI leads to alterations in brain function and pathology caused by a blow or other external force on the head [15]. As a major cause of disability and mortality, ABIs are increasingly being considered as a public health burden and place significant economic strain on society [16]. People who sustain an ABI can experience physical, cognitive, and communication impairments that are often long lasting and significantly impact their everyday functioning [3,17-20].

VR refers to "a computer-generated digital environment that can be experienced and interacted with as if that environment were real" [21]. VR systems are typically classified as immersive, semi-immersive, or nonimmersive [22,23], with immersion referring to the level of user perception with regard to being in a virtual environment (VE) rather than the real world [24]. Immersive VR systems supply VEs with a changing field of view via head-mounted displays (HMDs) [22,23,25]. Movement within immersive VEs is achieved via hardware such as head trackers, hand controllers, and body motion sensors [22,23]. Semi-immersive VR refers to systems that use projection-based systems (eg, driving simulators and use of shutter glasses), whereas nonimmersive VR systems include basic desktop displays and videogames [22,23,25]. To improve the delivery of assessment, treatment, and clinical outcomes for people with ABI, the use of VR should be further explored because of its potential to address limitations and produce benefits over conventional assessment and treatment methods [26-29].

Recommendations for VR Design and Implementation in ABI Rehabilitation

The existing literature provides guidance for safety and ethical considerations in clinical VR research [30-33], although there is a general lack of focus on design and development processes for using VR in ABI rehabilitation [34,35]. Some prior studies have proposed useful recommendations for implementing VR in health care and rehabilitation research; however, they are often specific to a particular VR system or do not focus on ABI [36-38]. Given the potentially limited applicability of these recommendations for designing and implementing VR in ABI rehabilitation, developing a set of recommendations would be beneficial for guiding research in this field.

Part 1 of this review had originally planned to include recommendations for using VR in TBI rehabilitation exclusively; however, no studies were identified. Examining the use of VR with other ABIs may provide guidance for this population. Recommendations specific to ABI are necessary, as individuals may experience motor, visual, or vestibular impairments that could impact their ability to use VR [39,40]. There are studies that propose recommendations for using VR in ABI rehabilitation [35,41]; however, there are no known systematic reviews on this topic. Developing recommendations based on existing studies and frameworks would be valuable for determining critical technological factors to guide the design and implementation of immersive VR in ABI rehabilitation.

VR for ABI Rehabilitation

ABI rehabilitation aims to improve function or provide compensatory strategies to reduce impairments and increase participation in activities and quality of life [42-45]. Goals and opportunities to practice real-life, meaningful tasks should be provided to maximize function and enable participation outside of clinical settings [43,46]. Examples include practicing cognitive, physical, or communication therapy goals in everyday activities [47,48] or relevant community settings (eg, home, cafes, and work) [49,50]. Early and intensive therapy is also recommended [51-53]. This focus on generalization and intensity can support neuroplastic changes, which in turn can assist with functional recovery following ABIs [51,54]. Furthermore, treatment programs should be goal oriented [53] and tailored to individual needs [55]; however, this can be challenging due to the complexity and diversity of ABIs [19,44]. Other limitations of the current assessment and treatment approaches in ABI rehabilitation include difficulties providing sufficient intensive therapy to allow for neuroplastic change or provision of services to patients residing in rural areas [44,56]. Advances in technology have provided opportunities for researchers to investigate the use of VR as a novel rehabilitation tool to overcome some of these barriers [29,56].

The benefits of VR for ABI rehabilitation include enhanced ecological validity, the ability to maintain experimental control over assessment and treatment standardization [28,57-59], and the option to cater to individual skills and goals by controlling task complexity [59,60]. VR can also provide relatively naturalistic VEs [28,61] for repeated practice of functional tasks such as activities of daily living [28,58,62], which may assist with generalizing targeted skills [63]. VR can also enhance patient motivation [61,64], which is necessary for neurorehabilitation, as repeated practice is required to achieve adequate treatment outcomes. Furthermore, VR may reduce

XSL•FO
RenderX

barriers to accessing rehabilitation services, such as affordability and geographical isolation [56].

The development of VR for ABI rehabilitation should incorporate co-design design principles [65], which is lacking in the current literature, especially for TBI. Co-design engages intended users (eg, patients and therapists) in the design and development of products, including VR [21]. People with ABI can experience complex and debilitating impairments. By determining their specific needs and capabilities, VR systems can be developed to meet patient and therapist needs, improve success in clinical practice, and maximize therapeutic outcomes [21,40,66,67].

With regard to using VR in TBI rehabilitation specifically, there are no known systematic reviews that examine the evidence base for using immersive VR to assess and treat any impairment sustained from TBIs. Experimental and review studies have mainly investigated VR for assessing or treating cognitive or motor impairments [34,56,57,68-72]. This current evidence base provides some support for using VR for TBI rehabilitation; however, the quality of the evidence is relatively low [34], and many studies include nonimmersive and semi-immersive systems [45,73-77] rather than focusing on immersive VR technology with HMDs. Issues identified in experimental studies include heterogeneity (eg, severity of TBI, VR system used, and outcome measures) with small sample sizes and a lack of randomized controlled trials (RCTs), resulting in the inability to perform meta-analysis [68,71]. Existing reviews provide important information; however, many reviews are not systematic in design or lack quality appraisal of included studies [34,45,56,57,70], focus on VR for only cognitive or motor impairments [68-72], do not consider immersive VR only [45,56,57,68-72,78], or do not review both assessment and treatment studies [56,57,68,69,72,78]. As literature to date has not focused on using immersive VR across the clinical spectrum of TBI rehabilitation, this review aims to identify and evaluate the use of immersive VR for the assessment and treatment of any impairment sustained within this group.

Objectives

This systematic review contains two parts and aims to:

1. Identify and synthesize existing recommendations and frameworks for designing and implementing therapeutic VR for ABI rehabilitation. By doing so, we aim to identify key technological and co-design factors to propose recommendations for the systematic development of VR apps in this field.
2. Determine the current published evidence base for using immersive VR for TBI assessment and treatment. The identified studies will be compared against the synthesized recommendations from part 1 to determine strengths and potential gaps in the literature with reference to recommendations for VR design and implementation to propose ways to improve future research and practice.

Methods

Protocol and Registration

This review has been registered with the International Prospective Register of Systematic Reviews (CRD42020152884) and was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement [79].

Search Strategy

A systematic search was conducted in August 2019. A total of 11 databases were accessed: CINAHL, Cochrane Central, Embase, Institute of Electrical and Electronics Engineers Xplore, MEDLINE, ProQuest Central, PsycBITE, PsychINFO, Scopus, speechBITE, and Web of Science. Search strategies were adapted for individual database requirements. Gray literature was also searched to ensure that all relevant studies were identified (ie, peer-reviewed conference proceedings and clinical guidelines). Additional studies were sourced by hand searching the reference lists of the included papers and repeating database searches in June 2020.

Two systematic searches were conducted to address the research aims in this review. For part 1, the general search strategy was *virtual reality* AND *assessment* OR *intervention* OR *research* AND *recommendations* AND *neurorehabilitation* OR *brain injury*. For part 2, the general search terms were *virtual reality* AND *traumatic brain injury* AND *assessment* OR *intervention*. The following limits were placed on the searches and studies for inclusion: (1) peer-reviewed, (2) full-text availability, (3) written in English, and (4) published between 2009 and 2020. Literature from the past 10 years was included to reflect current research, as the use of VR in health care and rehabilitation is changing rapidly. A detailed search strategy is provided in Multimedia Appendix 1, and the inclusion and exclusion criteria are provided in Textbox 1.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria for included studies.

Inclusion and Exclusion Criteria

Part 1

- Provided clear guidelines, consensus statements, recommendations, considerations, or pathways for using virtual reality with adults aged ≥18 years with an acquired brain injury, or the study referred to acquired brain injury populations.
- All study designs were considered.
- Included data or review of existing scientific evidence as a basis for recommendations.
- All virtual reality types were considered (recommendations for development and implementation are likely to be applicable for all therapeutic virtual reality designs despite potential variability in virtual reality systems and levels of immersion [66]).
- Papers that provided recommendations for a specific virtual reality system were excluded (ie, recommendations could not be applied to using virtual reality for acquired brain injury rehabilitation more broadly).

Part 2

- Included adults aged ≥18 years with a diagnosis of traumatic brain injury (studies were required to have ≥50% participants with a traumatic brain injury).
- Evaluated use of immersive virtual reality for assessing or treating any impairment sustained from a traumatic brain injury (immersive virtual reality was considered due to rapid advancements in technology [80]).
- Intervention studies included pre-post outcomes.
- Original research design (eg, randomized controlled trial, case series, and case study).
- Review papers and studies with semi-immersive or nonimmersive virtual reality systems were excluded.

Screening Process

The following process was conducted separately for each systematic search. Search results were exported to a reference manager (EndNote X9, Clarivate), where any duplicate references were excluded. Nonduplicate references were exported to a systematic review management program (Covidence) [81] to review titles and abstracts against search terms and eligibility criteria. An independent reviewer completed the reliability screening for 25.04% (426/1701) of randomly selected nonduplicate references from both searches. All disagreements were resolved via discussion. For part 1, the interrater point-by-point agreement was 97.6% (243/249), Cohen κ was 0.733, and 95% CI was 0.598-0.948, indicating substantial agreement [82]. For part 2, the interrater point-by-point agreement was 91.5% (162/177), Cohen κ was 0.605, and 95% CI was 0.426-0.785, indicating substantial agreement [82]. Full texts of eligible papers were retrieved and assessed for inclusion following this process.

Data Extraction

The following data were extracted and entered into a Microsoft Excel [83] spreadsheet by the first author. Part 1 included bibliographic data, study design and level of evidence, population, focus of the paper, and VR recommendations. Part 2 included bibliographic data, study design and level of evidence, participant characteristics, VR system and equipment, VR task, dosage and time in VR, outcome measures, results, adverse effects, inclusion and exclusion criteria, and evaluation against recommendations synthesized from studies in part 1.

Data Synthesis and Analysis

For part 1, data were synthesized via a qualitative meta-synthesis: (1) extracting recommendations from the included studies, (2) coding individual recommendations, (3) grouping recommendations based on similarities, and (4) synthesis of grouped recommendations to produce a single comprehensive list [84]. For part 2, meta-analysis was not performed due to the heterogeneity of the included studies. Therefore, the results are presented descriptively and in summary tables.

Quality Assessment

Where possible, studies were classified according to the Oxford Centre of Evidence-Based Medicine Levels of Evidence [85]. This hierarchy classifies evidence according to research methodology from the highest (level 1, systematic reviews) to the lowest (level 5, mechanism-based reasoning) levels of evidence. Methodological quality was assessed using appraisal tools relevant to study design [86]: A Measurement Tool to Assess Systematic Reviews for systematic reviews [87]; Joanna Briggs Institute (JBI) checklists for case series [88], case studies [88], qualitative studies [89], and papers of text and opinion (eg, consensus, expert opinion, and perspectives) [90]; and the Mixed Methods Appraisal Tool for mixed methods studies [91]. The PRISMA extension reporting guideline was used for scoping reviews [92], as no suitable appraisal tools were found. The first author and an independent rater conducted the quality assessment. Any disagreements were resolved through consensus with the fourth author. Studies were not excluded based on appraisal tool scores. The scores were used to guide the interpretation of the results.

XSL•FO
RenderX

Results

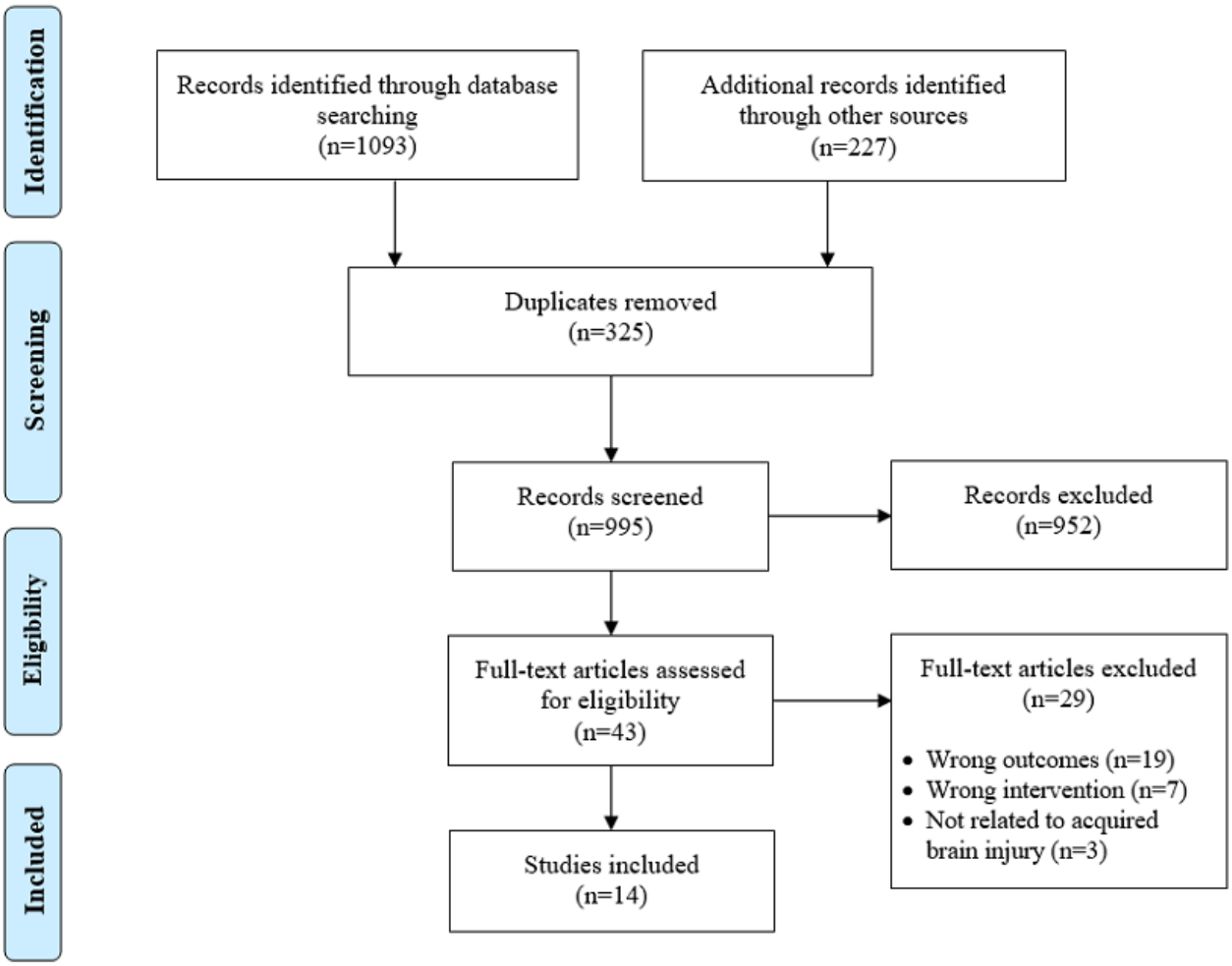
Part 1

Part 1 of this review aimed to develop recommendations for the design and implementation of therapeutic VR for ABI rehabilitation based on a synthesis of the existing literature.

Study Selection

Database, gray literature, and hand searches returned 1320 potential studies. Following the removal of duplicates, 995 studies were reviewed for keywords and eligibility criteria. After reading the full texts, 14 studies met the criteria for this review (Figure 1).

Figure 1. PRISMA flow diagram for studies included in part 1.



Study Design and Level of Evidence

A variety of study designs were included: 1 systematic review [10], 4 literature reviews [93-96], 5 text and opinion papers [39-41,66,67], 2 mixed methods studies [97,98], 1 scoping review [35], and 1 qualitative study [99]. A total of 5 of these studies were published as conference proceedings [39,93,96,98,99]. Further information regarding the details of the included studies is presented in Table 1.

XSL•FO
RenderX

Table 1. Study characteristics of the papers included in part 1.

Author	Country	Study design	Population or participant details	VR[a] definition (VR equipment and environment)	Aims of the study
Birckhead et al [66]	United States	Expert opinion consensus	Clinical health care and rehabilitation (makes references to stroke)	Immersive VR, defined VR as using an "HMD[b] with a close proximity screen"	To develop a methodological, best practice framework to guide development and implementation of high-quality therapeutic VR in health care
Bryant et al [99]	Australia	Qualitative[c]	Rehabilitation, including ABI[d] (communication disability)	Immersive VR	To explore views of health care and VR professionals on VR-based rehabilitation
Deutsch and Westcott McCoy [93]	United States	Literature review[c]	Neurological rehabilitation	Mentions nonimmersive, semi-immersive, and immersive VR systems	To review literature on VR in neurorehabilitation and offer suggestions for bridging gaps between research and practice when adopting VR
Glegg and Levac [39]	Canada	Perspective or discussion[c]	Rehabilitation (based on neurorehabilitation)	Mentions nonimmersive, semi-immersive, and immersive VR systems	To provide recommendations for the development, research, and clinical implementation of VR based on known barriers and facilitators
Glegg and Levac [35]	Canada	Scoping review	Rehabilitation (including ABI)	Included studies used a range of nonimmersive, semi-immersive, and immersive VR systems	To determine factors that contribute to facilitators and barriers to implementing VR in rehabilitation and to develop recommendations to address barriers
Kellmeyer [40]	Germany	Perspective or discussion	Neurology and psychiatry	Immersive VR	To discuss implications of using highly immersive VR systems within neurology and psychiatry, including ethical issues and adverse effects
Laver et al [10]	Australia	Systematic review (1[e])	Stroke	Included studies used a range of nonimmersive, semi-immersive, and immersive VR systems	To determine efficacy of VR for stroke rehabilitation
Lee et al [97]	Korea	Mixed methods	Acute stroke; 8 participants (4 male and 4 female; mean age 63 years)	Semi-immersive VR system (Microsoft Kinect; whack-a-mole game for upper limb movement)	To explore patients' perceived difficulty and enjoyment during VR rehabilitation and the factors affecting experiences and to suggest implementation strategies for VR-based rehabilitation for acute stroke
Levin et al [94]	Canada	Literature review	ABI (upper limb impairments)	Defines VR with examples of nonimmersive, semi-immersive, and immersive systems	To review motor control and learning principles and to discuss how they can be included in the design of VR training environments
Lewis and Rosie [95]	New Zealand	Literature review	Chronic neurological conditions (associated movement disorders)	Included studies used a range of nonimmersive, semi-immersive, and immersive VR systems	To review studies that examine users' responses to VR interventions and develop suggestions for how future VR systems can address user needs and expectations
Proffitt and Lange [41]	United States	Perspective or discussion	Stroke	VR systems that allow for immersion without assistance (ie, robotic devices)	To outline steps for developing VR interventions for stroke rehabilitation

XSL•FO
RenderX

Author	Country	Study design	Population or participant details	VR[a] definition (VR equipment and environment)	Aims of the study
Proffitt et al [67]	United States	Review of case studies	Rehabilitation (including stroke and TBI[f])	Included studies used a range of nonimmersive and semi-immersive VR systems	To review examples of end user involvement in VR research to provide recommendations for user-engaged design and implementation for VR in clinical practice
Ramírez-Fernández et al [96]	Mexico	Literature review[c]	Stroke (upper limb impairments)	Not specified; all VR environments	To develop a taxonomy of VR design factors for upper limb rehabilitation of stroke patients
Vaezipour et al [98]	Australia	Mixed methods[c]	Speech pathologists trialed a VR system designed for neurological conditions including ABI (communication impairments)	Immersive VR system (HTC VIVE Pro; simulated kitchen activity)	To explore speech pathologists' perspectives about immersive VR for rehabilitation of neurogenic communication disorders and to determine advantages and barriers to VR use

[a]VR: virtual reality.

[b]HMD: head-mounted display.

[c]Conference proceeding.

[d]ABI: acquired brain injury.

[e]Oxford levels of evidence (not applied to mixed methods or qualitative papers).

[f]TBI: traumatic brain injury.

Populations and Participants

Studies included a range of participants or populations: ABIs (8/14, 57%) [39,40,67,93-95,98,99], stroke only (4/14, 29%) [10,41,96,97], or a variety of medical conditions with reference to ABIs (2/14, 14%) [35,66].

VR Details

Various VR systems and levels of immersion were considered or described in the included studies: a combination of nonimmersive, semi-immersive, and immersive systems (7/14, 50%) [10,35,39,67,93-95]; immersive systems (4/14, 29%) [40,66,98,99]; semi-immersive systems (1/14, 7%) [97]; any VR environment (1/14, 7%) [96]; and VR systems that provided immersion without robotic devices (1/14, 7%) [41].

Qualitative Data Synthesis and Analysis

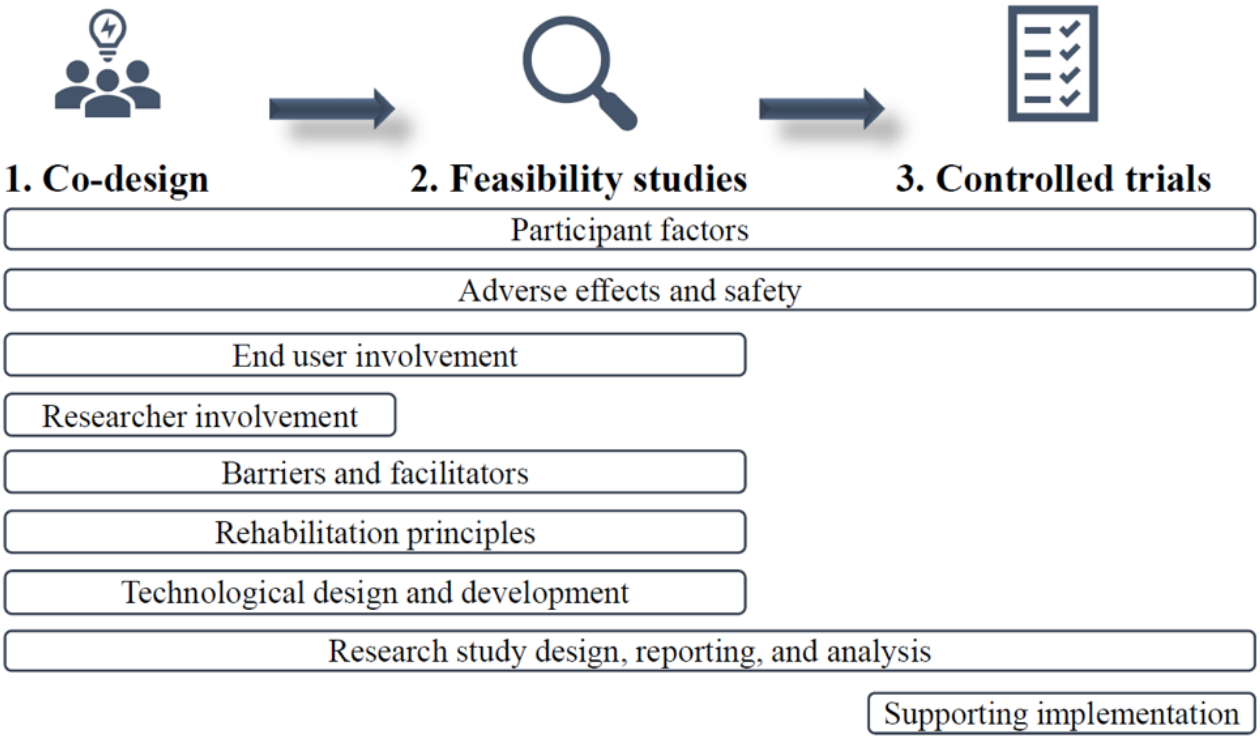
Overview

Three key phases of therapeutic VR development in health care were recommended according to the methodological framework developed by Birckhead et al [66]: (1) content development via end user involvement and iterative testing processes; (2) testing for feasibility, acceptability, tolerability, and efficacy; and (3) conducting RCTs for VR versus control interventions. Similar processes were described by Proffitt and Lange [41] and Laver et al [10], who recommended that initial VR studies for stroke rehabilitation should determine safety, validity, and usability with intended users before conducting larger trials and comparative studies [10,41].

Further synthesis and analysis of recommendations from the 14 included studies identified nine categories of recommendations related to participant, design, and technology factors for VR development and implementation in ABI rehabilitation: (1) end user involvement; (2) participant factors; (3) adverse effects and safety; (4) researcher involvement; (5) barriers and facilitators; (6) rehabilitation principles; (7) technological design and development; (8) supporting implementation; and (9) research study design, reporting, and analysis. Many of these categories are interlinked and can be considered across the suggested phases of design and implementation for therapeutic VR (Figure 2). A summary of the nine categories of recommendations is presented in Textbox 2, and a complete downloadable version is provided in Multimedia Appendix 2 [10,35,39-41,66,67,93-99].

XSL•FO
RenderX

Figure 2. Phases and categories of recommendations for virtual reality design and implementation.



XSL•FO
RenderX

Textbox 2. Summarized recommendations for design and implementation of virtual reality for acquired brain injury rehabilitation.

Category and Recommendations

End User Involvement

- Involve end users when designing virtual reality apps [39,40,66,67,95,98].

Participant Factors

- Consider participant factors when designing prototypes (eg, age, gender, ethnicity, health conditions, social position, cognition, and physical limitations) [40,66,95-98].
- Determine the impact of virtual reality on motivation and how to sustain engagement [10,66,98].
- Observe users to learn about their behavior [66].

Adverse Effects and Safety

- Measure and report physical and emotional adverse effects [40,66].
- Examine safety of virtual reality devices and tasks to determine suitability and contraindications [39,67,98].

Researcher Involvement

- Develop ideas and evaluate virtual reality prototypes as a team [66].

Determining Barriers and Facilitators to Virtual Reality

- Identify potential barriers and facilitators to designing and implementing virtual reality with key stakeholders [35,39,66,67] and offer solutions or implementation strategies [35].

Rehabilitation Principles

- Maintain therapeutic principles in virtual reality tasks (eg, principles of motor learning) [94,95].
- Tasks should be progressively challenging and customizable [94-98].
- When providing feedback, consider real-time knowledge of performance [94-96] and multimodal feedback (eg, visual, auditory, and haptic) [94,96,97].

Technological Design and Development

- Use hardware and software that is unrestrictive and allows for movement and possible postural constraints [95].
- Work in collaboration with virtual reality experts, game developers, and engineers [35,39,40].

Supporting Implementation

- Support therapists with virtual reality [35,39,67,93,99] and provide continued training [35,39,93].
- Provide information, training, and support for patients using virtual reality [95,96,98,99].

Research Study Design, Reporting, and Analysis

- Conduct larger, adequately powered trials [10].
- For randomized controlled trials, use appropriate randomization, conceal allocation, use CONSORT (Consolidated Standards of Reporting Trials) guidelines [66], and justify and describe control conditions [10,66,93].
- When reporting virtual reality research, consider using reporting guidelines (eg, Template for Intervention Description and Replication) [66], describe intervention details [66,67] and efforts to conceal allocation clearly [66], register trials on a publicly accessible registry, and publish all research regardless of outcomes [66].
- When selecting outcome measures, consider clinical relevance and validity [66,67], patient-reported outcomes [66,67], and pre- and postintervention measures [66,94]; measure long-term effects of virtual reality interventions [10,66]; compare against nonrandomized control groups [66]; and evaluate virtual reality in natural environments [93].

End User Involvement

Involving end users in co-design for therapeutic VR was recommended in 6 studies [39,40,66,67,95,98]. Suggested end users include therapists and patients who ultimately benefit from the VR systems. Co-design encourages those involved in the design process to gather user feedback to improve the iterations of VR tasks under development. It was recommended that this feedback includes, for example, patient willingness to try VR, what worked or did not work, or which VR systems therapists wish to acquire. Gathering user feedback, both positive and negative, is an important part of co-design because VR prototypes should be iteratively tested by end users and

XSL•FO
RenderX

continually refined to better meet patient and therapist needs [66,67].

Participant Factors

A range of participant factors should be considered when developing therapeutic VR for ABI rehabilitation [10,40,66,95,96,98]. This is because individuals with neurological impairments may experience physical and cognitive conditions that could impact their understanding and use of VR. In addition, more meaningful and effective VR programs can be developed when specific user needs are considered [40,66,95,97]. Observing patient behavior in a clinical context or conducting surveys and interviews may provide insights into these factors to be considered [66]. Ways to enhance and sustain patient motivation should also be addressed [10,66,98].

Adverse Effects and Safety

Participants may experience a range of potential adverse physical and emotional effects when using VR. Some of the potential adverse effects of using VR include headaches, vertigo, nausea, dizziness, fear, and anxiety [66]. A total of 5 of the included studies [39-41,66,98] recommended that these adverse effects should be measured and reported during the development of VR tasks for neurological populations. Measuring adverse effects is necessary to establish a research base for the safety of VR devices and programs and to determine any contraindications for use with people who have an ABI [39-41,98].

Determining Barriers and Facilitators

Potential barriers and facilitators of VR use and implementation should be identified via site-specific assessments or interviews during VR development [35,39,66,67]. These barriers and facilitators should be from the perspective of patients and health care providers. Solutions to address the identified barriers should be provided to support successful VR design and implementation [35]. This is particularly important during the design and feasibility testing of therapeutic VR [66].

Researcher Involvement in Design

Birckhead et al [66] provided recommendations for researcher involvement and collaboration in the initial design processes. Research teams should develop several ideas for VR tasks and then determine the most feasible and suitable ideas for prototype testing. It is argued that team collaboration is essential for developing therapeutic VR and can often lead to more innovative and improved designs [66].

Rehabilitation Principles

A total of 5 studies [94-98] provided recommendations for incorporating rehabilitation principles (eg, principles of motor learning and control [94,100]) when designing VR tasks for ABI rehabilitation. Motor patterns used during VR tasks should provide patients with rehabilitation benefits [95], and tasks should be able to be modified to accommodate impairment severities and stages of recovery [94-98]. Recommendations were also made to provide feedback. Some studies suggested that feedback on performance should be given in real time to engage and motivate patients [94-96]. Levin et al [94] recommended that knowledge of both performance and results should be provided in a way that does not interrupt task progression. Multimodal feedback (eg, visual, auditory, and haptic) should also be considered to potentially improve engagement [94,96]. Additional recommendations included designing tasks that have purposeful goals and providing the opportunity for multiple repetitions of rehabilitation targets [94].

Technological Design and Development

A total of 4 of the included studies [35,39,40,95] discussed technological factors to be considered when designing therapeutic VR. Researchers and therapists are recommended to work with game developers and engineers, as they have technological and design expertise to build tasks that meet patients' and therapists' needs [35,39,40]. Hardware (eg, HMDs and hand-held devices) and software for VR tasks should also be carefully considered to prevent potential limitations or failures of the VR technology [95]. Examples include designing or selecting systems that allow for adequate movement, providing a sufficient field of view, and reducing the complexity of hardware.

Supporting Implementation

Recommendations for supporting the implementation of VR in practice for therapists and patients were provided in 8 of the included studies [35,39,67,93,95,96,98,99]. Recommendations for therapists included providing tailored clinical training packages, providing education about using VR to achieve rehabilitation outcomes, and identifying ways to assist with troubleshooting and implementation [35,39,67,93,99]. Education and training could also be provided to students in relevant professions [39,93]. In terms of supporting patients, therapists should provide adequate information about the purpose of VR and clear instructions for use [98,99] and monitor patient performance regardless of practice settings (eg, rehabilitation units, home-based therapy), as it is necessary for rehabilitation tasks to be performed correctly to achieve sufficient treatment outcomes [95,96].

Study Design, Reporting, and Analysis

Among the included studies, 4 [10,41,66,93] proposed recommendations for the design, analysis, and reporting of clinical VR research. Recommendations were related, but not limited, to ensuring rigorous randomization processes in clinical trials, use of reporting guidelines, and detailing specific components of the VR tasks (eg, equipment, therapy dose, and intensity). Recommendations were provided for selecting and using outcome measures to determine the effectiveness of VR interventions [10,41,66,93]. Researchers are encouraged to use outcome measures that are clinically relevant, validated, and standardized, and researchers should also consider using patient-reported measures [66] or target participation outcomes [93]. With regard to the timing of outcome measures, Birckhead et al [66] recommended that measures should be taken at least pre- and postintervention, whereas Laver et al [10] suggested that outcome measures should be taken at least 3 months postintervention to determine long-term effects.

XSL•FO
RenderX

Methodological Quality

The methodological quality of the included studies varied (Multimedia Appendix 3 [10,35,39-41,66,67,87-92,97-99,101-103]). Cross-comparison of study quality could not be made, as quality assessment could not be conducted with a single appraisal tool and no suitable appraisal tools were found for all study types. Studies with relatively high methodological quality included one systematic review [10], papers of text and opinion [39-41,66,67], one qualitative focus group study [99], one scoping review [35], and one mixed methods study [97]. The second mixed methods study [98] met four out of five criteria for quantitative methodology but had limitations in reporting qualitative and mixed methods components. However, this study was presented as a conference proceeding, so all details may not have been included.

TBI Guidelines

Published guidelines for the management of TBI were included as gray literature in the search for this review. None of the reviewed guidelines [47,48,50,53,104-110] provided clear recommendations for the development or implementation of VR for TBI rehabilitation. Despite this, three guidelines suggested that VR is a priority area for research for the assessment and management of TBI [53,104,105], highlighting the potential of VR for TBI rehabilitation and the importance of conducting research in this area.

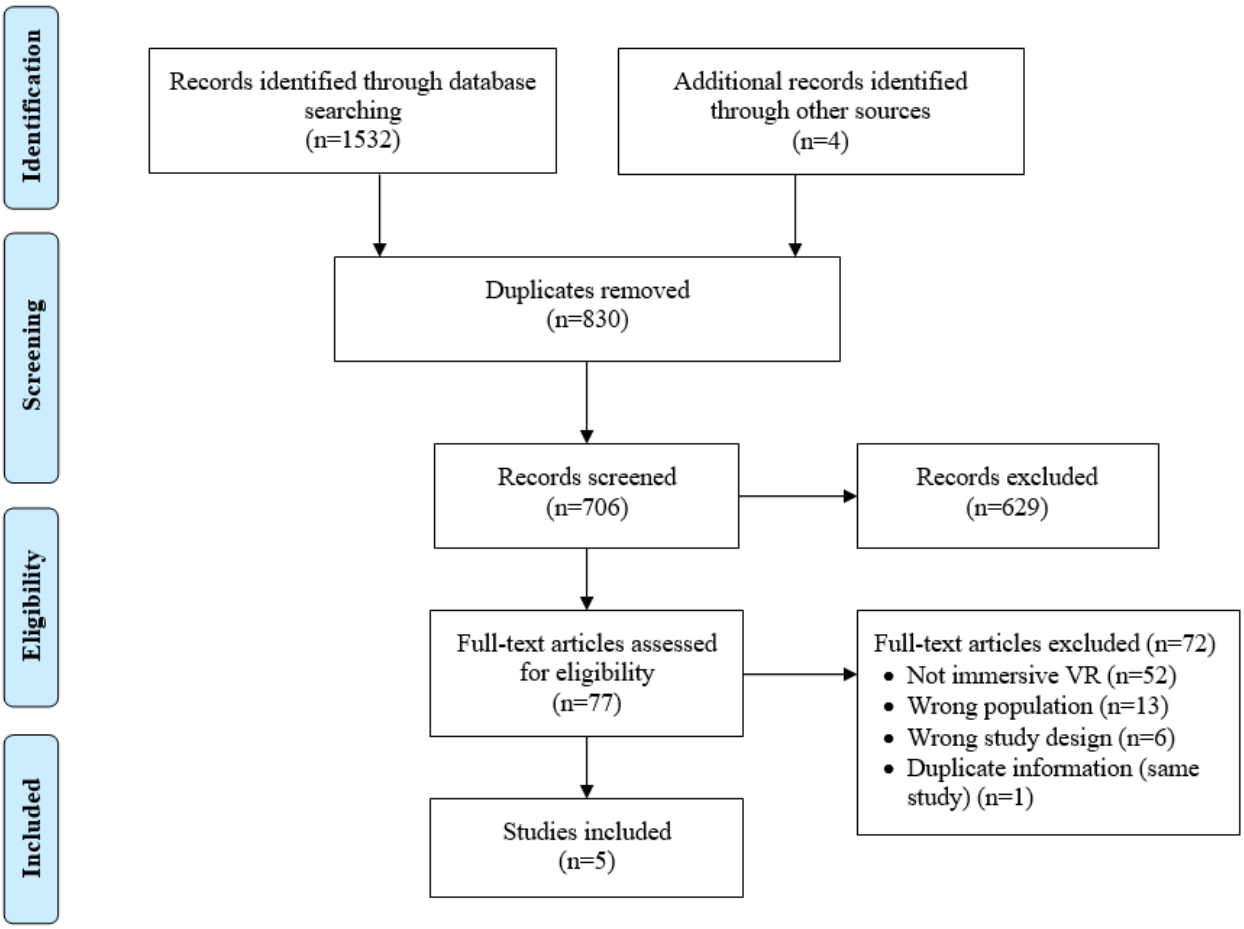
Part 2

Part 2 of this review aimed to identify current evidence for using immersive VR for assessment and treatment in TBI rehabilitation. These studies were also examined to determine the extent to which they incorporate recommendations for developing and implementing therapeutic VR based on the findings from part 1 of this review.

Study Selection

Database, gray literature, and hand searches returned 1536 potential studies. A total of 830 duplicate studies were removed. Following the screening of titles and abstracts, 77 nonduplicates were identified for full-text screening. Of these studies, 5 met the inclusion criteria. This process is illustrated in Figure 3.

Figure 3. PRISMA flow diagram for studies included in part 2. VR: virtual reality.



Study Design and Level of Evidence

Included studies investigated the use of VR for assessment (2/5, 40%) [111,112] or treatment (3/5, 60%) [101-103] of impairments following TBI (Table 2). The overall rating of the level of evidence [85] was low. Assessment studies provided level 4 evidence [111,112]. Treatment study designs included 1 case series (level 4) [101] and 2 single case studies [102,103].

XSL•FO
RenderX

Table 2. Study characteristics and participant details of the studies included in part 2.

Author	Country	Study design	Participant numbers (TBI[a] severity)	Age (years)	Gender	Time post TBI	Setting
Banville and Nolin [111]	Canada	Quasi-experimental assessment (4[b])	TBI=31 (7 moderate and 24 severe) and matched healthy controls=31	TBI: mean 27 (SD 11) and controls: mean 27 (SD 11)	TBI: 23 males and 8 females; controls: 23 males and 8 females	Mean 3.78 (SD 2.5) years	Outpatient
Cikajlo et al [101]	Slovenia	Case series (4[b])	TBI=3, brain tumor=1, and nonbrain injury=4	TBI or brain tumor: range 24-48 and nonbrain injury: range 27-40	Not reported	Not reported	Outpatient
Gamito et al [102]	Portugal	Case study	TBI=1 (severe)	20	Male	3 months	Inpatient rehabilitation ward
Ma et al [103]	United States	Case study[c]	TBI=1 (moderate-severe)	26	Male	9 months	Physical therapy clinic
Robitaille et al [112]	Canada	Proof of concept (4[b])	TBI=6 (mild) and healthy controls=6	TBI: mean 30.3 (SD 8.6; range 18-61) and controls: mean 30.3 (SD 5.3)	Not reported	Median 0.46 years; range 2 weeks to 7 years	Not reported

[a]TBI: traumatic brain injury.

[b]Oxford levels of evidence.

[c]Conference proceeding.

Participant Characteristics

A total of 42 participants with TBI were included in this study (Table 2). The number of participants ranged from 1 to 31 (mean 8, SD 13). Time post injury ranged from 2 weeks to 7 years. Although not always reported, the majority of participants were males (25/42, 60%) aged between 18 and 61 years. Most of the included participants sustained a severe TBI (25/42, 60%) [102,111], followed by moderate TBI (7/42, 17%) [111], mild TBI (6/42, 14%) [112], and moderate-severe TBI (1/42, 2%) [103]. TBI severity was not reported for 3 participants [101]. Where reported, VR was used in inpatient [102] and outpatient settings [101,103,111].

Target of VR Assessment or Treatment

Impairments targeted in VR assessment included executive functions [112] and prospective memory [111]. VR treatments targeted attention and working memory [102], balance and functional mobility [103], and stress and anxiety [101].

VR Details

Keeping with the definition of immersive VR systems, all studies used HMDs to create immersive VEs [101-103,111,112], and one study included body motion tracking [112]. Virtual cities were used as the basis for memory and attention tasks [102,111]. Other VR tasks included a mindfulness-based stress reduction program [101], a military patrol task to assess executive functions [112], and standing balance practice in VR as an adjunct to traditional physical therapy [103].

Assessment studies did not report the time spent in VR. Where reported, therapy session duration ranged from 5 to 25 minutes; total dosage ranged from 50 minutes to 3 hours; and participants received 5, 8, or 10 therapy sessions. One study provided breaks during VR sessions [103]. Further information related to the VR details of the included studies is presented in Table 3.

XSL•FO
RenderX

Table 3. Virtual reality details of the studies included in part 2.

Study	Target	Dosage or time in VR[a] and VR hardware	Task details	Outcome measures	Results	Adverse effects and potential issues	Eligibility criteria
Banville and Nolin [111]	Prospective memory and executive functions	Time in VR not reported; HMD[b] (eMagin Z800) with head tracker	Non-VR task: prospective memory assessment based on Rivermead Behavioral Memory Test; VR task: virtual prospective memory tasks completed in a virtual city (included visiting apartments and selecting an apartment to live in)	Non-VR: correct actions, time to complete, and whether prompting was required; VR: prospective memory score, precision score, time to complete, success in task, and IPQ[c]	Participants could be classified as having a TBI[d] by performance on each task. TBI participants were significantly less precise with prospective memory VR tasks (P=.02) and took more time to perform VR tasks (P=.008) than controls	SSQ[e] completed; no reported cybersickness; SSQ scores did not differ between groups	Inclusion criteria: confirmed TBI
Cikajlo et al [101]	Stress and anxiety	8 sessions (25 minutes per session, once weekly); Samsung Galaxy X7 mobile phone mounted to HMD (Samsung Gear VR)	Mindfulness stress reduction program conducted by an instructor (eg, self-meditation and group discussions in various VEs[f] such as a mountain view or a room with a fireplace)	MAAS,[g] SWLS,[h] MMSE[i] (TBI only), session and task ratings, and head motion	Slight improvement in MAAS and SWLS scores (TBI group>non-TBI group); one participant increased MMSE (not reported for others); task ratings: simple to use and interesting; varying head motions	Potential for overheating of mobile phones (sessions were, therefore, limited to 30 minutes)	Inclusion criteria: mild or no cognitive impairment and able to understand instructions; exclusion criteria: high diopters, astigmatism, and wore glasses
Gamito et al [102]	Working memory and attention	10 sessions (5 minutes each session); HMD (eMagin Z800)	Activities included performing ADLs[j] in the VE (eg, breakfast, navigating to and from a supermarket, and buying items)	PASAT[k] and completion time of each task	Significant increase in correct responses between initial and final PASAT scores (P<.05)	Not reported	Inclusion criteria: diagnosed with a TBI 3-12 months prior, clinical deficit in memory and attention, and aged 18-60 years; exclusion criteria: a previous psychiatric disorder that may impact memory and attention and neurological diseases
Ma et al [103]	Balance and functional mobility	5 sessions (12 trials, with a 1- to 2-minute break in between); breaks were decided by the participant. Samsung Galaxy X7 mobile phone mounted to HMD (Samsung Gear VR)	Standing balance exercises in a VE with traffic lights, street crossing and traffic island, night and day versions, moving cars, and static buildings	DGI,[l] mini-BEST,[m] DHI,[n] ABC,[o] GROC,[p] and patient-specific functional scale (self-scoring street crossing and multitasking abilities)	Improvements in DGI, mini-BEST, DHI, GROC, and patient-specific functional scale	Not reported	Not reported

XSL•FO
RenderX

Study	Target	Dosage or time in VR[a] and VR hardware	Task details	Outcome measures	Results	Adverse effects and potential issues	Eligibility criteria
Robitaille et al [112]	Executive functions	Time in VR not reported; HMD (piSight 166-43) with head tracking and body tracking (MoCap)	Exploration of a simulated military patrol scene in a village with different conditions and obstacles to navigate (eg, fences, wires, beams, and avatars)	PQ,[q] SUS,[r] errors, walking speed and fluidity, and obstacle clearance	TBI group walked faster and had slightly greater obstacle clearances. Significant difference in walking fluidity between groups for two hostile blocks (P=.046). Moderate to high presence scores on SUS and PQ	SSQ completed: 1 participant with TBI and 1 control participant reported slight headaches	Inclusion criteria (TBI): mild TBI; inclusion criteria (controls): no known TBI or other neurological or musculoskeletal issues

[a]VR: virtual reality.
[b]HMD: head-mounted display.
[c]IPQ: Igroup Presence Questionnaire.
[d]TBI: traumatic brain injury.
[e]SSQ: Simulator Sickness Questionnaire.
[f]VE: virtual environment.
[g]MAAS: Mindfulness Attention Awareness Scale.
[h]SWLS: Satisfaction With Life Scale.
[i]MMSE: Mini-Mental State Examination.
[j]ADL: activities of daily living.
[k]PASAT: Paced Auditory Serial Addition Task.
[l]DGI: Dynamic Gait Index.
[m]mini-BEST: Mini-Balance Evaluation System Test.
[n]DHI: Dizziness Handicap Index.
[o]ABC: Activities Balance Confidence Scale.
[p]GROC: Global Rating of Change.
[q]PQ: Presence Questionnaire.
[r]SUS: Slater-Usoh-Steed Questionnaire.

Comparisons

One assessment study investigated participant performance on a VR and non-VR assessment of prospective memory [111], with participants performing similarly on both tasks. Both assessment studies [111,112] included healthy control groups, with findings able to distinguish between participants with and without TBI. One intervention study [101] included unmatched participants without TBI and did not report major differences in outcomes between the groups.

Eligibility Criteria and Adverse Effects

One intervention study [101] excluded participants due to visual impairments that may have impacted their tolerance of the HMD. The Simulator Sickness Questionnaire [113] was used in 2 studies [111,112] to monitor potential adverse effects of VR, with only 1 participant with TBI reporting a slight headache [112]. A total of 2 studies did not consider or report adverse effects [102,103].

Outcome Measures and Results

Various outcome measures were used, and all studies included more than one measure. VR task-specific outcome measures were used in 4 studies [101,102,111,112]. A total of 4 studies used outcome measures traditionally used for non-VR tasks [101-103,111]. Depending on the outcome, intervention studies took measures pre-, mid-, and postintervention [101,102] or pre-post intervention with a 1-month follow-up [103].

The results from the assessment studies suggested that VR assessment tasks have the potential for use as novel diagnostic tools [111,112]. These studies included healthy controls and could classify participants as having a TBI or not by their performance on VR assessment tasks. The findings demonstrated a significant difference between groups for walking fluidity during a navigation task [112] and for time and precision to complete a procedural memory task in VR [111].

Statistically significant outcomes were reported in one case study [102], where VR intervention for attention and memory deficits led to a significant increase in pre-post assessment scores on the Paced Auditory Serial Addition Task. The remaining treatment studies presented descriptive data on the outcomes of participants with TBI [101,103]. Ma et al [103] demonstrated that VR treatment combined with standard physical therapy led to improvements in gait and balance. Cikajlo et al [101] reported slight improvements in pre-post psychometric outcomes for stress and anxiety following the implementation of a

XSL•FO
RenderX

group-based VR mindfulness intervention that participants rated as interesting and simple to use.

Methodological Quality

The quality assessments of the included treatment studies are presented in Multimedia Appendix 3. We assessed 2 case studies using the JBI checklist for case reports [88] and these studies were found to have low [102] to moderate [103] methodological quality. The studies included adequate information about interventions but did not provide comprehensive participant or assessment details or refer to measuring the potential adverse effects of VR. One study [101] was appraised using the JBI checklist for case series [88] and had a low methodological quality. Information about condition measurement and treatment outcomes was provided, but important details about the participants and recruitment methods were absent.

Recommendations for VR: Evidence in TBI Studies

VR assessment and treatment studies for TBI rehabilitation were examined regarding the three suggested phases of VR development [66] and the nine categories of recommendations for VR design and implementation proposed in part 1 of this review (Table 4). With regard to phases of VR design, there was one feasibility study [101] and one proof-of-concept study [112] but no controlled trials or co-design studies with detailed descriptions of end user involvement in VR development. A total of 3 studies [101,111,112] have considered the potential adverse effects of VR. Rehabilitation principles were included in studies that provided varied or progressively challenging tasks [102,103]. All treatment studies collected at least one outcome measure pre- and postintervention and included clinically relevant outcome measures. At least one patient-reported outcome measure was included in 4 studies [101,103,111,112], and one study included a user survey for task feedback [101]. Details were generally provided about the *active ingredients* of the VR equipment and tasks (eg, dose, repetitions, and time in VR).

Recommendations that were not included in the assessment and treatment studies were as follows: involving researchers when developing VR tasks, considering barriers and facilitators to VR use, technological design and development, and supporting VR in practice. However, many of these recommendations are applicable to specific phases of VR development and implementation, so they may not have been relevant for all studies. Furthermore, recommendations may have been addressed but not specifically reported on.

Table 4. Inclusion of recommendations for virtual reality design and implementation in traumatic brain injury studies.

Recommendations for VR[a] development	Banville and Nolin [111]	Cikajlo et al [101]	Gamito et al [102]	Ma et al [103]	Robitaille et al [112]
Phase of VR development					
Phase 1: co-design					
Phase 2: feasibility		✓[b]			✓
Phase 3: controlled trials					
Recommendation					
End user involvement		✓			
Participant factors					
Adverse effects and safety	✓	✓			✓
Researcher involvement					
Determining barriers and facilitators to VR					
Rehabilitation principles			✓	✓	
Technological design and development					
Supporting implementation					
Research study design, reporting, and analysis	✓	✓	✓	✓	✓

[a]VR: virtual reality.

[b]Recommendation present.

Discussion

Principal Findings

Overview

The findings of this systematic review highlight that research in the field of VR and ABI rehabilitation, particularly for TBI, is still emerging. To our knowledge, this is the first study to synthesize existing recommendations for developing VR for ABI rehabilitation and to systematically review the current evidence base for using immersive VR for TBI rehabilitation. Recommendations for future research have been provided based on the results of this review.

XSL•FO
RenderX

Part 1

Part 1 of this review aimed to identify and synthesize the recommendations for designing and implementing therapeutic VR for ABI rehabilitation, with a focus on using existing frameworks to determine key technological and co-design factors. The findings appear to be consistent across VR technologies and health care settings and contain important considerations for using VR with people who have an ABI.

Three phases for VR development and implementation of therapeutic VR in health care were developed by Birckhead et al [66] and formed a framework against which this review was completed. A total of nine categories of recommendations were subsequently developed from all 14 studies included in part 1 of this review. Most recommendations that were addressed in the limited literature reported in this study were related to the design of VR tasks, including consideration of participant factors, involving key end users and researchers, determining barriers and facilitators of VR use, technological considerations, and including rehabilitation principles in VR tasks. The recommendations can be applied throughout VR development and implementation (Figure 2).

A phased approach to VR design should be considered [10,41,66], with early focus on engaging key end users [35,39,40,45,66,93,98,114] in co-design and feasibility studies before conducting larger controlled trials [10,41,66,115]. This approach is not widely used in research investigating VR for health care purposes [115-117] or for adults with TBI, yet user involvement in VR development is emerging in pediatric TBI [118] and ABI rehabilitation [98,99]. Involving end users in designing digital health interventions is recommended [66,114] and is essential for producing successful VR apps [40,66,67], particularly for people with TBI [119]. Additional recommendations for research design, reporting and analysis, and supporting implementation for VR were synthesized to further guide and strengthen research in this area. Such recommendations include the use of reporting guidelines such as the Template for Intervention Description and Replication [120] and the CONSORT (Consolidated Standards of Reporting Trials) [121] and supporting end users with VR adoption via education and training.

The included studies drew on research examining various VR systems and levels of immersion. This reflects the literature from the past decade and highlights the limited use of fully immersive VR for neurological rehabilitation. However, recommendations for VR research are similar across VR platforms, particularly for design and feasibility studies [66], and may be adapted for various clinical settings and disciplines [67]. No specific recommendations were made regarding preferred VR hardware or software. Future research and clinical VR apps should focus on more immersive systems [40] because of the rapid advancement and availability of VR technology as well as tasks that can be used across different devices to facilitate transition and use from rehabilitation facilities to home environments (eg, wireless systems and mobile phone compatibility) [39,122,123]. The tolerance and safety of new VR systems will need to be established for people with ABI [39-41,66,98].

Part 2

The second part of this review aimed to determine the current evidence base for using immersive VR for TBI rehabilitation and to review the extent to which these studies addressed the recommendations developed in part 1 of this review. A total of 5 studies that investigated the use of immersive VR for TBI assessment and treatment were identified and included.

The findings demonstrate a small body of evidence for using immersive VR in TBI rehabilitation. Studies have used immersive VR to assess cognitive impairment following mild, moderate, and severe TBI [111,112]. VR treatments targeted memory and attention [102] and balance [103] in single cases of participants with moderate-severe TBI and anxiety in a case series of 3 people with TBI (severity not disclosed) [101]. The range of time post injury, age of participants, clinical settings, dosage and frequency, and VR tasks suggest that immersive VR has potential for use with people with TBI across the continuum of care [68]; however, further studies are required to support this evidence because of the limited number of included studies and small sample sizes.

Three different HMDs were used in the 5 studies, including the smartphone-compatible Samsung Gear, which highlights the accessibility and affordability of immersive VR technology [80,115,123]. There were no commonalities in terms of VR software and tasks, with each study implementing VR tasks specific to the targeted impairments and outcomes. There would be potential challenges in developing a VR platform to suit the various impairments and severities of brain injury [95], so the proposed recommendations developed in part 1 of this review may improve the consistency of VR development for TBI rehabilitation where heterogeneity may not be accounted for.

There were limited adverse effects of VR use reported in 3 of the 5 included studies [101,111,112]. The potential adverse effects of VR use must be monitored and reported due to the limited research in this area and to determine the safety and feasibility of immersive VR for people with TBI [39-41,66]. Involving VR experts and interdisciplinary teams should be considered when designing new VR tasks to mitigate potential safety issues [32].

The included studies reported positive findings, but few specific conclusions can be drawn regarding assessment and treatment effectiveness due to a limited number of studies with small sample sizes, a lack of control conditions, assessment reference standards, face-to-face comparisons, and heterogeneity of data that prevented pooling of data and meta-analysis. Some studies had relatively low methodological quality and provided minimal details about participants and recruitment methods, making it challenging to generalize findings and determine the suitability of VR platforms and tasks for people with TBI.

The current evidence base for using immersive VR for TBI rehabilitation incorporates some of the recommendations proposed in part 1 of this review (Table 4), yet there is a need for future studies to increase end user engagement in co-design and feasibility testing before conducting controlled trials [10,41,66]. Specifically, future research should engage end users [39,40,66,67,95] and clinical experts alongside the design and

XSL•FO
RenderX

technology industries to inform VR development [35,39,95], identify potential barriers and facilitators to using VR [35,39,66,67], and focus on stepwise progression of VR research [10,41,66,114]. By doing so, VR tasks for TBI, and ABI rehabilitation more broadly, can better meet patients' and therapists' needs [35,39,40,66,67].

Study Limitations

Although a systematic literature search was undertaken, some existing studies may have been excluded, as inclusion criteria limited papers to English only, and gray literature did not include conference abstracts or theses. In addition, inconsistencies with VR definitions and classifications [124-127] may have led to exclusion based on the definitions of VR immersion levels. Caution should be taken when interpreting results, as the overall levels of evidence presented were relatively low. Furthermore, the inclusion of various study designs led to the inability to use a single critical appraisal tool, and some of the included studies presented poor methodological quality.

There were limited high-level evidence studies that provided recommendations for developing and implementing VR in ABI rehabilitation (ie, part 1). Although this may decrease the perceived value of findings, it likely reflects the fact that VR technology and practice in this field are still emerging [35] and may be moving faster than the evidence base [66,93]. Despite this, most recommendations were synthesized from a VR expert consensus paper [66], which provided a framework for therapeutic VR methodology. Furthermore, findings across the included studies were similar and provide a basis for ongoing research for developing and implementing VR for ABI rehabilitation, including for people with TBI.

The current evidence base for using immersive VR for TBI assessment and treatment (ie, part 2) consisted mainly of lower-quality methods of case studies and case series. These study designs may be suitable for early co-design and feasibility studies for VR development, yet this was not always reflected in the included studies. On the basis of the methodological quality and levels of evidence, future studies should provide important details about participants, recruitment methods, and interventions; consider and report on adverse effects; and include reference standards and control conditions. These findings reflect the general lack of high-quality evidence, as highlighted in previous reviews of nonimmersive, semi-immersive, and immersive VR for TBI rehabilitation [34,56,68,70,71], and the findings may be difficult to generalize due to heterogeneity. The use of the proposed recommendations may improve the consistency of design and implementation of VR in TBI rehabilitation and provide a model to advance clinical research in this area.

Recommendations for Future Research

Future research should consider the proposed recommendations when designing and implementing VR tasks for ABI rehabilitation, especially for people with TBIs. As identified in this review, stepwise VR development (Figure 2) [10,41,66] is lacking in current TBI literature. There needs to be an increased focus on co-design processes to investigate the opinions and needs of key end users, including people with ABI and their therapists [10,35,39,41,66,67,114]. Iterative testing and feasibility studies will also be necessary to establish the safety and viability of new immersive VR tasks before implementing larger-scale studies and RCTs [10,41,66], particularly given that people with ABIs may face challenges when using and interacting with VR systems [39,40] and there are limited studies that use immersive VR for TBI rehabilitation [101-103,111,112].

Although this review offers a starting point for guiding future research in VR for TBI rehabilitation, the recommendations provided were formed from papers that included a wider range of ABIs. Work should be undertaken to develop guidelines specific to TBI to ensure more rigorous development and evaluation of therapeutic VR for this population. Expanding the evidence base for using VR with people with TBI has been encouraged and highlighted as a priority area in published guidelines for TBI management [53,104,105]. Furthermore, immersive VR to date has been used to treat people with TBI who have cognitive disorders, balance issues, or anxiety. Future research should develop and investigate the use of VR for other significant impairments that people with TBI may experience, such as cognitive-communication disorders.

Conclusions

This systematic review highlights that the use of immersive VR in ABI rehabilitation, especially TBI, is still in its infancy. There are no existing guidelines for designing and implementing VR tasks specific to TBI, reflecting the need for more rigorous research in this area. Existing evidence demonstrates the potential to use immersive VR for TBI assessment and treatment. However, this comprises a small number of lower-quality studies with a large degree of heterogeneity, small sample sizes, and limited generalizability of the findings.

This review produced recommendations for developing and implementing VR for ABI rehabilitation (Textbox 2 and Multimedia Appendix 2): engaging end users; considering participant, researcher, and technological factors; addressing facilitators and barriers; incorporating rehabilitation principles; and supporting implementation in clinical practice. These recommendations can be incorporated into the three phases of therapeutic VR development (Figure 2). Researchers in ABI rehabilitation are presented with an opportunity to capitalize on the current digital health movement, particularly when the required technology and resources for immersive VR are becoming increasingly available and affordable. VR has the potential to provide innovative assessment and treatment methods, and future work in this field should use these recommendations to improve consistency, quality, and outcomes for the effective design of therapeutic VR.

XSL•FO
RenderX

Acknowledgments

This research was supported by a Research Training Program Scholarship and a Merit Award Scholarship awarded to the first author (SB). The authors would like to thank Petra Avramovic for assistance with the reliability screening and critical appraisal of the included papers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example database searches.
[PDF File (Adobe PDF File), 185 KB - jmir_v23i7e26344_app1.pdf]

Multimedia Appendix 2
List of recommendations for the design and implementation of virtual reality for acquired brain injury rehabilitation.
[PDF File (Adobe PDF File), 142 KB - jmir_v23i7e26344_app2.pdf]

Multimedia Appendix 3
Quality appraisal of the included studies.
[XLSX File (Microsoft Excel File), 23 KB - jmir_v23i7e26344_app3.xlsx]

References

1. Adamovich SV, Fluet GG, Tunik E, Merians AS. Sensorimotor training in virtual reality: a review. NeuroRehabilitation 2009;25(1):29-44 [FREE Full text] [doi: 10.3233/NRE-2009-0497] [Medline: 19713617]
2. Bohil CJ, Alicea B, Biocca F. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 3;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
3. Teasell R, Bayona N, Marshall S, Cullen N, Bayley M, Chundamala J, et al. A systematic review of the rehabilitation of moderate to severe acquired brain injuries. Brain Inj 2007 Feb;21(2):107-112. [doi: 10.1080/02699050701201524] [Medline: 17364527]
4. What Is the Difference Between Acquired Brain Injury and Traumatic Brain Injury? Brain Injury Association of America. 2020. URL: https://www.biausa.org/brain-injury/about-brain-injury/nbiic/what-is-the-difference-between-an-acquired-brain-injury-and-a-traumatic-brain-injury [accessed 2020-11-10]
5. Connelly L, Jia Y, Toro M, Stoykov ME, Kenyon RV, Kamper DG. A pneumatic glove and immersive virtual reality environment for hand rehabilitative training after stroke. IEEE Trans. Neural Syst. Rehabil. Eng 2010 Oct;18(5):551-559 [FREE Full text] [doi: 10.1109/tnsre.2010.2047588]
6. Kizony R, Katz N, (Tamar) Weiss PL. Adapting an immersive virtual reality system for rehabilitation. J Visual Comput Animat 2003 Dec;14(5):261-268 [FREE Full text] [doi: 10.1002/vis.323]
7. Lee SH, Jung H, Yun SJ, Oh B, Seo HG. Upper extremity rehabilitation using fully immersive virtual reality games with a head mount display: a feasibility study. PM R 2020 Mar;12(3):257-262. [doi: 10.1002/pmrj.12206] [Medline: 31218794]
8. Levin MF, Magdalon EC, Michaelsen SM, Quevedo AA. Quality of grasping and the role of haptics in a 3-d immersive virtual reality environment in individuals with stroke. IEEE Trans Neural Syst Rehabil Eng 2015 Nov;23(6):1047-1055. [doi: 10.1109/tnsre.2014.2387412]
9. McNeill M, Pokluda L, McDonough S, Crosbie C. Immersive Virtual Reality for Upper Limb Rehabilitation Following Stroke. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. 2004 Oct Presented at: IEEE'04; October 10-13, 2004; The Hague, Netherlands URL: https://doi.org/10.1109/ICSMC.2004.1400754 [doi: 10.1109/ICSMC.2004.1400754]
10. Laver KE, Lange B, George S, Deutsch JE, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
11. Jaffe DL, Brown DA, Pierson-Carey CD, Buckley EL, Lew HL. Stepping over obstacles to improve walking in individuals with poststroke hemiplegia. J Rehabil Res Dev 2004 May;41(3A):283-292 [FREE Full text] [doi: 10.1682/jrrd.2004.03.0283] [Medline: 15543446]
12. Crosbie JH, Lennon S, McGoldrick MC, McNeill MD, McDonough SM. Virtual reality in the rehabilitation of the arm after hemiplegic stroke: a randomized controlled pilot study. Clin Rehabil 2012 Sep;26(9):798-806. [doi: 10.1177/0269215511434575] [Medline: 22275463]
13. Jannink MJ, Erren-Wolters CV, de Kort AC, van der Kooij H. An electric scooter simulation program for training the driving skills of stroke patients with mobility problems: a pilot study. Cyberpsychol Behav 2008 Dec;11(6):751-754. [doi: 10.1089/cpb.2007.0271] [Medline: 18954289]

XSL•FO
RenderX

14. Cortés-Pérez I, Nieto-Escamez FA, Obrero-Gaitán E. Immersive virtual reality in stroke patients as a new approach for reducing postural disabilities and falls risk: a case series. Brain Sci 2020 May 15;10(5):296 [FREE Full text] [doi: 10.3390/brainsci10050296] [Medline: 32429085]
15. Menon DK, Schwab K, Wright DW, Maas AI, Demographics and Clinical Assessment Working Group of the InternationalInteragency Initiative toward Common Data Elements for Research on Traumatic Brain InjuryPsychological Health. Position statement: definition of traumatic brain injury. Arch Phys Med Rehabil 2010 Nov;91(11):1637-1640. [doi: 10.1016/j.apmr.2010.05.017] [Medline: 21044706]
16. Humphreys I, Wood RL, Phillips CJ, Macey S. The costs of traumatic brain injury: a literature review. Clinicoecon Outcomes Res 2013;5:281-287 [FREE Full text] [doi: 10.2147/CEOR.S44625] [Medline: 23836998]
17. Boyle GJ, Haines S. Severe traumatic brain injury: some effects on family caregivers. Psychol Rep 2002 Apr;90(2):415-425. [doi: 10.2466/pr0.2002.90.2.415] [Medline: 12061578]
18. Cullen N. Canadian healthcare perspective in traumatic brain injury rehabilitation. J Head Trauma Rehabil 2007;22(4):214-220. [doi: 10.1097/01.HTR.0000281836.54602.6d] [Medline: 17667063]
19. Turner-Stokes L, Pick A, Nair A, Disler PB, Wade DT. Multi-disciplinary rehabilitation for acquired brain injury in adults of working age. Cochrane Database Syst Rev 2015 Dec 22(12):CD004170. [doi: 10.1002/14651858.CD004170.pub3] [Medline: 26694853]
20. Menon DK, Bryant C. Time for change in acquired brain injury. Lancet Neurol 2019 Jan;18(1):28. [doi: 10.1016/S1474-4422(18)30463-0] [Medline: 30563679]
21. Jerald J. The VR Book: Human-Centred Design for Virtual Reality. New York, NY: Association for Computing Machinery and Morgan & Claypool; 2016.
22. Ma M, Zheng H. Virtual reality and serious games in healthcare. In: Brahnam S, Jain LC, editors. Advanced Computational Intelligence Paradigms in Healthcare 6. Virtual Reality in Psychotherapy, Rehabilitation, and Assessment. Heidelberg, Germany: Springer, Berlin, Heidelberg; 2011:169-192.
23. Kalawsky R. Developments in VR Technology. The Advisory Group on Computer Graphics. 1996. URL: http://www.agocg.ac.uk/reports/virtual/vrtech/title.htm [accessed 2019-06-09]
24. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3549-3557 [FREE Full text] [doi: 10.1098/rstb.2009.0138] [Medline: 19884149]
25. Miller HL, Bugnariu NL. Level of immersion in virtual environments impacts the ability to assess and teach social skills in autism spectrum disorder. Cyberpsychol Behav Soc Netw 2016 Apr;19(4):246-256 [FREE Full text] [doi: 10.1089/cyber.2014.0682] [Medline: 26919157]
26. Burdea GC. Virtual rehabilitation--benefits and challenges. Methods Inf Med 2003;42(5):519-523. [Medline: 14654886]
27. Rizzo A. Virtual reality and disability: emergence and challenge. Disab Rehab 2009 Jul 7;24(11-12):567-569. [doi: 10.1080/09638280110111315]
28. Rizzo A, Kim GJ. A swot analysis of the field of virtual reality rehabilitation and therapy. Presence: Teleoperators Virtual Environ 2005;14(2):119-146. [doi: 10.1162/1054746053967094]
29. Massetti T, da Silva TD, Crocetta TB, Guarnieri R, de Freitas BL, Bianchi Lopes P, et al. The clinical utility of virtual reality in neurorehabilitation: a systematic review. J Cent Nerv Syst Dis 2018;10:1179573518813541 [FREE Full text] [doi: 10.1177/1179573518813541] [Medline: 30515028]
30. Madary M, Metzinger TK. Recommendations for good scientific practice and the consumers of VR-technology. Front Robot AI 2016 Feb 19;3:3 [FREE Full text] [doi: 10.3389/frobt.2016.00003]
31. Ramirez E. Ecological and ethical issues in virtual reality research: a call for increased scrutiny. Philosophical Psychology 2018 Oct 13;32(2):211-233 [FREE Full text] [doi: 10.1080/09515089.2018.1532073]
32. Bryant L, Brunner M, Hemsley B. A review of virtual reality technologies in the field of communication disability: implications for practice and research. Disabil Rehabil Assist Technol 2020 May;15(4):365-372. [doi: 10.1080/17483107.2018.1549276] [Medline: 30638092]
33. Spiegel JS. The ethics of virtual reality technology: social hazards and public policy recommendations. Sci Eng Ethics 2018 Oct;24(5):1537-1550. [doi: 10.1007/s11948-017-9979-y] [Medline: 28942536]
34. Aida J, Chau B, Dunn J. Immersive virtual reality in traumatic brain injury rehabilitation: a literature review. NeuroRehabilitation 2018;42(4):441-448. [doi: 10.3233/NRE-172361] [Medline: 29660958]
35. Glegg SM, Levac DE. Barriers, facilitators and interventions to support virtual reality implementation in rehabilitation: a scoping review. PM R 2018 Nov;10(11):1237-51.e1 [FREE Full text] [doi: 10.1016/j.pmrj.2018.07.004] [Medline: 30503231]
36. Campelo A, Hashim J, Weisberg A, Katz L. Virtual Rehabilitation in the elderly: Benefits, issues, and considerations. In: 2017 International Conference on Virtual Rehabilitation. 2017 Presented at: VR'17; June 19-22, 2017; Montreal, Canada p. 19-22. [doi: 10.1109/icvr.2017.8007485]
37. Kitson A, Prpa M, Riecke BE. Immersive interactive technologies for positive change: a scoping review and design considerations. Front Psychol 2018;9:1354 [FREE Full text] [doi: 10.3389/fpsyg.2018.01354] [Medline: 30123161]
38. Lindner P, Miloff A, Hamilton W, Reuterskiöld L, Andersson G, Powers MB, et al. Creating state of the art, next-generation Virtual Reality exposure therapies for anxiety disorders using consumer hardware platforms: design considerations and future directions. Cogn Behav Ther 2017 Sep;46(5):404-420. [doi: 10.1080/16506073.2017.1280843] [Medline: 28270059]

XSL•FO
RenderX

39. Glegg SM, Levac DE. Enhancing Clinical Implementation of Virtual Reality: an Evidence-based Perspective and Call to Action. In: 2017 International Conference on Virtual Rehabilitation. 2017 Presented at: VR'17; June 19-22, 2017; Montreal, Canada. [doi: 10.1109/icvr.2017.8007488]
40. Kellmeyer P. Neurophilosophical and ethical aspects of virtual reality therapy in neurology and psychiatry. Camb Q Healthc Ethics 2018 Oct;27(4):610-627. [doi: 10.1017/S0963180118000129] [Medline: 30198465]
41. Proffitt R, Lange B. Considerations in the efficacy and effectiveness of virtual reality interventions for stroke rehabilitation: moving the field forward. Phys Ther 2015 Mar;95(3):441-448 [FREE Full text] [doi: 10.2522/ptj.20130571] [Medline: 25343960]
42. Report to Congress on Traumatic Brain Injury in the United States: Epidemiology and Rehabilitation. Centers for Disease Control and Prevention. 2015. URL: https://www.cdc.gov/traumaticbraininjury/pdf/TBI_Report_to_Congress_Epi_and_Rehab-a.pdf [accessed 2020-02-03]
43. Carlson PM, Boudreau ML, Davis J, Johnston J, Lemsky C, McColl MA, et al. 'Participate to learn': a promising practice for community ABI rehabilitation. Brain Inj 2006 Oct;20(11):1111-1117. [doi: 10.1080/02699050600955337] [Medline: 17123927]
44. Maas AI, Menon DK, Adelson PD, Andelic N, Bell MJ, Belli A, InTBIR ParticipantsInvestigators. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 2017 Dec;16(12):987-1048. [doi: 10.1016/S1474-4422(17)30371-X] [Medline: 29122524]
45. Zanier ER, Zoerle T, Di Lernia D, Riva G. Virtual reality for traumatic brain injury. Front Neurol 2018;9:345 [FREE Full text] [doi: 10.3389/fneur.2018.00345] [Medline: 29867748]
46. International Classification of Functioning, Disability and Health. World Health Organisation. 2001. URL: https://apps.who.int/iris/handle/10665/42407 [accessed 2019-08-09]
47. Tate R, Kennedy M, Ponsford J, Douglas J, Velikonja D, Bayley M, et al. INCOG recommendations for management of cognition following traumatic brain injury, part III: executive function and self-awareness. J Head Trauma Rehabil 2014;29(4):338-352. [doi: 10.1097/HTR.0000000000000068] [Medline: 24984096]
48. Togher L, Wiseman-Hakes C, Douglas J, Stergiou-Kita M, Ponsford J, Teasell R, INCOG Expert Panel. INCOG recommendations for management of cognition following traumatic brain injury, part IV: cognitive communication. J Head Trauma Rehabil 2014;29(4):353-368. [doi: 10.1097/HTR.0000000000000071] [Medline: 24984097]
49. Mattingly EO. Functional evaluation and treatment in acquired brain injury acute rehabilitation. Perspect ASHA SIGs 2018 Jan;3(2):13-20 [FREE Full text] [doi: 10.1044/persp3.sig2.13]
50. Bayley MT, Tate R, Douglas JM, Turkstra LS, Ponsford J, Stergiou-Kita M, INCOG Expert Panel. INCOG guidelines for cognitive rehabilitation following traumatic brain injury: methods and overview. J Head Trauma Rehabil 2014;29(4):290-306. [doi: 10.1097/HTR.0000000000000070] [Medline: 24984093]
51. Turner-Stokes L. Evidence for the effectiveness of multi-disciplinary rehabilitation following acquired brain injury: a synthesis of two systematic approaches. J Rehabil Med 2008 Oct;40(9):691-701 [FREE Full text] [doi: 10.2340/16501977-0265] [Medline: 18843419]
52. Hylin MJ, Kerr AL, Holden R. Understanding the mechanisms of recovery and/or compensation following injury. Neural Plast 2017;2017:7125057 [FREE Full text] [doi: 10.1155/2017/7125057] [Medline: 28512585]
53. Section II: Assessment and Rehabilitation of Brain Injury Sequelae. Ontario Neurotrauma Foundation. 2016. URL: https://braininjuryguidelines.org/modtosevere/fileadmin/Guidelines_components/Rec/Section_2_REC_complete_ENG_final.pdf [accessed 2019-07-08]
54. Königs M, Beurskens EA, Snoep L, Scherder EJ, Oosterlaan J. Effects of timing and intensity of neurorehabilitation on functional outcome after traumatic brain injury: a systematic review and meta-analysis. Arch Phys Med Rehabil 2018 Jun;99(6):1149-59.e1. [doi: 10.1016/j.apmr.2018.01.013] [Medline: 29428344]
55. Khan F, Baguley IJ, Cameron ID. 4: rehabilitation after traumatic brain injury. Med J Aust 2003 Mar 17;178(6):290-295. [doi: 10.5694/j.1326-5377.2003.tb05199.x] [Medline: 12633489]
56. Pietrzak E, Pullman S, McGuire A. Using virtual reality and videogames for traumatic brain injury rehabilitation: a structured literature review. Games Health J 2014 Aug;3(4):202-214. [doi: 10.1089/g4h.2014.0013] [Medline: 26192369]
57. Edwards J, Vess J, Reger G, Cernich A. The use of virtual reality in the military's assessment of service members with traumatic brain injury: recent developments and emerging opportunities. Appl Neuropsychol Adult 2014;21(3):220-230. [doi: 10.1080/09084282.2013.796554] [Medline: 25084846]
58. Martínez-Moreno JM, Sánchez-González P, Luna M, Roig T, Tormos JM, Gómez EJ. Modelling ecological cognitive rehabilitation therapies for building virtual environments in brain injury. Methods Inf Med 2016;55(1):50-59. [doi: 10.3414/ME15-01-0050] [Medline: 26391897]
59. Schultheis M, Himelstein J, Rizzo A. Virtual reality and neuropsychology: upgrading the current tools. J Head Trauma Rehabil 2002 Oct;17(5):378-394. [doi: 10.1097/00001199-200210000-00002] [Medline: 12802250]
60. Weiss P, Katz N. The potential of virtual reality for rehabilitation. J Rehabil Res Dev 2004 Sep;41(5):vii-vix [FREE Full text] [Medline: 15558392]
61. Holden MK. Virtual environments for motor rehabilitation: review. Cyberpsychol Behav 2005 Jun;8(3):187-211; discussion 212. [doi: 10.1089/cpb.2005.8.187] [Medline: 15971970]

XSL•FO
RenderX

62. Besnard J, Richard P, Banville F, Nolin P, Aubin G, le Gall D, et al. Virtual reality and neuropsychological assessment: the reliability of a virtual kitchen to assess daily-life activities in victims of traumatic brain injury. Appl Neuropsychol Adult 2016;23(3):223-235. [doi: 10.1080/23279095.2015.1048514] [Medline: 26569498]
63. Brundage SB, Graap K, Gibbons KF, Ferrer M, Brooks J. Frequency of stuttering during challenging and supportive virtual reality job interviews. J Fluency Disord 2006;31(4):325-339. [doi: 10.1016/j.jfludis.2006.08.003] [Medline: 16999990]
64. Thornton M, Marshall S, McComas J, Finestone H, McCormick A, Sveistrup H. Benefits of activity and virtual reality based balance exercise programmes for adults with traumatic brain injury: perceptions of participants and their caregivers. Brain Inj 2005 Nov;19(12):989-1000. [doi: 10.1080/02699050500109944] [Medline: 16263641]
65. Norman D, Draper S. User-centred System Design: New Perspective on Human-computer Interaction. Hillsdale, NJ: Lawrence Earlbaum Associate; 1986.
66. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
67. Proffitt R, Glegg S, Levac D, Lange B. End-user involvement in rehabilitation virtual reality implementation research. J Enabling Technol 2019;13(2):92-100 [FREE Full text] [doi: 10.1108/JET-10-2018-0050] [Medline: 31663039]
68. Manivannan S, Al-Amri M, Postans M, Westacott LJ, Gray W, Zaben M. The effectiveness of virtual reality interventions for improvement of neurocognitive performance after traumatic brain injury: a systematic review. J Head Trauma Rehabil 2019;34(2):E52-E65. [doi: 10.1097/HTR.0000000000000412] [Medline: 30045223]
69. Alashram AR, Annino G, Padua E, Romagnoli C, Mercuri NB. Cognitive rehabilitation post traumatic brain injury: a systematic review for emerging use of virtual reality technology. J Clin Neurosci 2019 Aug;66:209-219. [doi: 10.1016/j.jocn.2019.04.026] [Medline: 31085075]
70. Maggio MG, De Luca R, Molonia F, Porcari B, Destro M, Casella C, et al. Cognitive rehabilitation in patients with traumatic brain injury: a narrative review on the emerging use of virtual reality. J Clin Neurosci 2019 Mar;61:1-4. [doi: 10.1016/j.jocn.2018.12.020] [Medline: 30616874]
71. Shin H, Kim K. Virtual reality for cognitive rehabilitation after brain injury: a systematic review. J Phys Ther Sci 2015 Sep;27(9):2999-3002 [FREE Full text] [doi: 10.1589/jpts.27.2999] [Medline: 26504344]
72. Larson EB, Feigon M, Gagliardo P, Dvorkin AY. Virtual reality and cognitive rehabilitation: a review of current outcome research. NeuroRehabilitation 2014;34(4):759-772. [doi: 10.3233/NRE-141078] [Medline: 24820166]
73. Mumford N, Duckworth J, Thomas PR, Shum D, Williams G, Wilson PH. Upper-limb virtual rehabilitation for traumatic brain injury: a preliminary within-group evaluation of the elements system. Brain Inj 2012;26(2):166-176. [doi: 10.3109/02699052.2011.648706] [Medline: 22360522]
74. Grealy MA, Johnson DA, Rushton SK. Improving cognitive function after brain injury: the use of exercise and virtual reality. Arch Phys Med Rehabil 1999 Jun;80(6):661-667. [doi: 10.1016/s0003-9993(99)90169-7] [Medline: 10378492]
75. Dvorkin AY, Ramaiya M, Larson EB, Zollman FS, Hsu N, Pacini S, et al. A "virtually minimal" visuo-haptic training of attention in severe traumatic brain injury. J Neuroeng Rehabil 2013 Aug 9;10:92 [FREE Full text] [doi: 10.1186/1743-0003-10-92] [Medline: 23938101]
76. Rábago CA, Wilken JM. Application of a mild traumatic brain injury rehabilitation program in a virtual realty environment: a case study. J Neurol Phys Ther 2011 Dec;35(4):185-193. [doi: 10.1097/NPT.0b013e318235d7e6] [Medline: 22027473]
77. Jacoby M, Averbuch S, Sacher Y, Katz N, Weiss PL, Kizony R. Effectiveness of executive functions training within a virtual supermarket for adults with traumatic brain injury: a pilot study. IEEE Trans Neural Syst Rehabil Eng 2013 Mar;21(2):182-190. [doi: 10.1109/TNSRE.2012.2235184] [Medline: 23292820]
78. Aulisio MC, Han DY, Glueck AC. Virtual reality gaming as a neurorehabilitation tool for brain injuries in adults: a systematic review. Brain Inj 2020 Aug 23;34(10):1322-1330. [doi: 10.1080/02699052.2020.1802779] [Medline: 32791020]
79. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
80. Snoswell AJ, Snoswell CL. Immersive virtual reality in health care: systematic review of technology and disease states. JMIR Biomed Eng 2019 Sep 26;4(1):e15025. [doi: 10.2196/15025]
81. Veritas Health Innovation: Melbourne, Australia. Covidence Systematic Review Software. URL: http://www.covidence.org/ [accessed 2019-02-07]
82. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
83. Microsoft Excel. Microsoft Corporation. 2020. URL: https://office.microsoft.com/excel [accessed 2020-02-02]
84. McArthur A, Klugárová J, Yan H, Florescu S. Innovations in the systematic review of text and opinion. Int J Evid Based Healthc 2015 Sep;13(3):188-195. [doi: 10.1097/XEB.0000000000000060] [Medline: 26207851]
85. The Oxford 2011 Levels of Evidence. OCEBM Levels of Evidence Working Group. URL: http://www.cebm.net/index.aspx?o=5653 [accessed 2019-09-10]
86. Buccheri RK, Sharifi C. Critical appraisal tools and reporting guidelines for evidence-based practice. Worldviews Evid Based Nurs 2017 Dec 12;14(6):463-472. [doi: 10.1111/wvn.12258] [Medline: 28898556]

XSL•FO
RenderX

87. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Br Med J 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
88. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, et al. Systematic reviews of etiology and risk. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. New York, USA: JBI; 2020.
89. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
90. McArthur A, Klugárová J, Yan H, Florescu S. Innovations in the systematic review of text and opinion. Int J Evid Based Healthc 2015 Sep;13(3):188-195. [doi: 10.1097/XEB.0000000000000060] [Medline: 26207851]
91. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
92. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 2;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
93. Deutsch JE, Westcott McCoy S. Prisma extension for scoping reviews (virtual reality and serious games in neurorehabilitation of children and adults: prevention, plasticity, and participation): checklist and explanation. Pediatr Phys Ther 2017 Jul;29(Suppl 3):S23-S36 [FREE Full text] [doi: 10.1097/PEP.0000000000000387] [Medline: 28654475]
94. Levin MF, Weiss PL, Keshner EA. Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles. Phys Ther 2015 Mar;95(3):415-425 [FREE Full text] [doi: 10.2522/ptj.20130579] [Medline: 25212522]
95. Lewis GN, Rosie JA. Virtual reality games for movement rehabilitation in neurological conditions: how do we meet the needs and expectations of the users? Disabil Rehabil 2012;34(22):1880-1886. [doi: 10.3109/09638288.2012.670036] [Medline: 22480353]
96. Ramírez-Fernández C, Morán AL, García-Canseco E, Orihuela-Espina F. Design Factors of Virtual Environments for Upper Limb Motor Rehabilitation of Stroke Patients. In: Proceedings of the 5th Mexican Conference on Human-Computer Interaction. 2014 Presented at: HCI'14; November 3-5, 2014; Oaxaca, Mexico. [doi: 10.1145/2676690.2676700]
97. Lee M, Pyun S, Chung J, Kim J, Eun S, Yoon B. A further step to develop patient-friendly implementation strategies for virtual reality-based rehabilitation in patients with acute stroke. Phys Ther 2016 Oct;96(10):1554-1564. [doi: 10.2522/ptj.20150271] [Medline: 27149961]
98. Vaezipour A, Aldridge D, Wall K, Koenig S, Theodoros T. Design and Development of a Virtual Reality System for the Management of Adult Neurogenic Communication Disorders. In: IEEE International Symposium on Mixed and Augmented Reality Adjunct. 2019 Presented at: IEEE'19; October 10-19, 2019; Beijing, China. [doi: 10.1109/ismar-adjunct.2019.00043]
99. Bryant L, Hemsley B, Bailey B, Bluff A, Nguyen V, Stubbs P, et al. Opportunities for Immersive Virtual Reality in Rehabilitation: Focus on Communication Disability. In: 53rd Hawaii International Conference on System Sciences. 2020 Presented at: CSS'20; January 7-10, 2020; Maui, HI. [doi: 10.24251/HICSS.2020.437]
100. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008 Feb;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
101. Cikajlo I, Cizman Staba U, Vrhovac S, Larkin F, Roddy M. A cloud-based virtual reality app for a novel telemindfulness service: rationale, design and feasibility evaluation. JMIR Res Protoc 2017 Jun 5;6(6):e108 [FREE Full text] [doi: 10.2196/resprot.6849] [Medline: 28583904]
102. Gamito P, Pacheo J, Morais D, Saraiva T, Lacerda R, Baptista A, et al. Traumatic brain injury memory training: a virtual reality online solution. Int J Disabil Hum Dev 2011;10(4):309-312 [FREE Full text] [doi: 10.1515/ijdhd.2011.049]
103. Ma L, Tornetta Jr F, Egan W, Wright WG. Virtual Reality Therapy as Adjunct to Traditional Physical Therapy for a TBI Patient Who Suffered a Gunshot Wound to the Head: Case Report. In: 13th International Conference on Virtual Rehabilitation. 2019 Presented at: VR'19; July 21-24, 2019; Tel Aviv, Israel. [doi: 10.1109/ICVR46560.2019.8994509]
104. Brain Injury Rehabilitation in Adults. Scottish Intercollegiate Guidelines Network (SIGN). 2013 Mar. URL: https://www.sign.ac.uk/media/1068/sign130.pdf [accessed 2019-08-07]
105. Marshall S, Janzen S, Duffy T, Faltynek P, Mirkowski M, Janssen S, et al. Motor and Sensory Impairment Rehabilitation Post Acquired Brain Injury. Evidence-Based Review of Moderate to Severe Acquired Brain Injury. 2018. URL: https://erabi.ca/wp-content/uploads/2018/12/Module4_V12_motorsensory.pdf [accessed 2019-08-08]
106. Occupational Therapy Practice Guidelines for Adults with Traumatic Brain Injury. National Guideline Clearinghouse (NGC). 2016. URL: http://www.guideline.gov/content.aspx?id=15287 [accessed 2019-08-08]
107. The Diagnosis, Acute Management and Rehabilitation of People with Traumatic Brain Injury. New Zealand Guidelines Group. 2016. URL: https://www.moh.govt.nz/NoteBook/nbbooks.nsf/0/B8738C3605889A6ACC257A6D00809243/$file/060831_TBI_Evidence_Tables_Final.pdf [accessed 2019-09-08]

XSL•FO
RenderX

108. Ponsford J, Bayley M, Wiseman-Hakes C, Togher L, Velikonja D, McIntyre A, INCOG Expert Panel. INCOG recommendations for management of cognition following traumatic brain injury, part II: attention and information processing speed. J Head Trauma Rehabil 2014;29(4):321-337. [doi: 10.1097/HTR.0000000000000072] [Medline: 24984095]
109. Velikonja D, Tate R, Ponsford J, McIntyre A, Janzen S, Bayley M, INCOG Expert Panel. INCOG recommendations for management of cognition following traumatic brain injury, part V: memory. J Head Trauma Rehabil 2014;29(4):369-386. [doi: 10.1097/HTR.0000000000000069] [Medline: 24984098]
110. Turkstra L, Ylvisaker M, Coelho C, Kennedy M, Sohlberg MM, Avery J, et al. Practice guidelines for standardized assessment for persons with traumatic brain injury. J Med Speech-Lang Path 2005;13(2):ix-xxxviii [FREE Full text]
111. Banville F, Nolin P. Using virtual reality to assess prospective memory and executive functions after traumatic brain injury. J Cyber Rehab 2012;5(1):45-55 [FREE Full text]
112. Robitaille N, Jackson PL, Hébert LJ, Mercier C, Bouyer LJ, Fecteau S, et al. A Virtual Reality avatar interaction (VRai) platform to assess residual executive dysfunction in active military personnel with previous mild traumatic brain injury: proof of concept. Disabil Rehabil Assist Technol 2017 Oct;12(7):758-764. [doi: 10.1080/17483107.2016.1229048] [Medline: 27677827]
113. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
114. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. Ideas (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
115. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
116. Huygelier H, Schraepen B, van Ee R, Vanden Abeele V, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. Sci Rep 2019 Mar 14;9(1):4519 [FREE Full text] [doi: 10.1038/s41598-019-41200-6] [Medline: 30872760]
117. Miller KJ, Adair BS, Pearce AJ, Said CM, Ozanne E, Morris MM. Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. Age Ageing 2014 Mar;43(2):188-195. [doi: 10.1093/ageing/aft194] [Medline: 24351549]
118. Shen J, Xiang H, Luna J, Grishchenko A, Patterson J, Strouse RV, et al. Virtual reality-based executive function rehabilitation system for children with traumatic brain injury: design and usability study. JMIR Serious Games 2020 Aug 25;8(3):e16947 [FREE Full text] [doi: 10.2196/16947] [Medline: 32447275]
119. Vaezipour A, Whelan B, Wall K, Theodoros D. Acceptance of rehabilitation technology in adults with moderate to severe traumatic brain injury, their caregivers, and healthcare professionals: a systematic review. J Head Trauma Rehabil 2019;34(4):E67-E82. [doi: 10.1097/HTR.0000000000000462] [Medline: 30608310]
120. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 7;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
121. Schulz K, Altman D, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Br Med J 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
122. Bohil CJ, Alicea B, Biocca FA. Virtual reality in neuroscience research and therapy. Nat Rev Neurosci 2011 Nov 3;12(12):752-762. [doi: 10.1038/nrn3122] [Medline: 22048061]
123. Jerdan SW, Grindle M, van Woerden HC, Kamel Boulos MN. Head-mounted virtual reality and mental health: critical review of current research. JMIR Serious Games 2018 Jul 6;6(3):e14 [FREE Full text] [doi: 10.2196/games.9226] [Medline: 29980500]
124. Cant R, Cooper S, Sussex R, Bogossian F. What's in a name? Clarifying the nomenclature of virtual simulation. Clinical Simulation in Nursing 2019 Feb;27:26-30. [doi: 10.1016/j.ecns.2018.11.003]
125. Kardong-Edgren S, Farra SL, Alinier G, Young HM. A call to unify definitions of virtual reality. Clin Simulat Nurs 2019 Jun;31:28-34. [doi: 10.1016/j.ecns.2019.02.006]
126. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence-Teleop Virt 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
127. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]

Abbreviations

ABI: acquired brain injury
CONSORT: Consolidated Standards of Reporting Trials
HMD: head-mounted display
JBI: Joanna Briggs Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

RCT: randomized controlled trial
TBI: traumatic brain injury
VE: virtual environment
VR: virtual reality

Edited by R Kukafka; submitted 09.12.20; peer-reviewed by A Vaezipour, J Salisbury, B Smith; comments to author 03.03.21; revised version received 25.04.21; accepted 24.05.21; published 30.07.21.

Please cite as:
Brassel S, Power E, Campbell A, Brunner M, Togher L
Recommendations for the Design and Implementation of Virtual Reality for Acquired Brain Injury Rehabilitation: Systematic Review
J Med Internet Res 2021;23(7):e26344
URL: https://www.jmir.org/2021/7/e26344
doi:10.2196/26344
PMID:34328434

©Sophie Brassel, Emma Power, Andrew Campbell, Melissa Brunner, Leanne Togher. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Unique Internet Search Strategies of Individuals With Self-Stated Autism: Quantitative Analysis of Search Engine Users' Investigative Behaviors

Eldad Yechiam[1*], PhD; Elad Yom-Tov[1,2*], PhD

[1]Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa, Israel

[2]Microsoft Research, Herzliya, Israel

[*]all authors contributed equally

Corresponding Author:
Eldad Yechiam, PhD
Faculty of Industrial Engineering and Management
Technion - Israel Institute of Technology
Technion City
Haifa
Israel
Phone: 972 48294420
Email: yeldad@ie.technion.ac.il

Abstract

Background: Although autism is often characterized in literature by the presence of repetitive behavior, in structured decision tasks, individuals with autism spectrum disorder (ASD) have been found to examine more options in a given time period than controls.

Objective: We aimed to examine whether this investigative tendency emerges in information searches conducted via the internet.

Methods: In total, 1746 search engine users stated that they had ASD in 2019. This group's naturally occurring responses following 1491 unique general queries and 78 image queries were compared to those of all other users of the search engine. The main dependent measure was scrolled distance, which denoted the extent to which additional results were scanned beyond the initial results presented on-screen. Additionally, we examined the number of clicks on search results as an indicator of the degree of search outcome exploitation and assessed whether there was a trade-off between increased search range and the time invested in viewing initial search results.

Results: After issuing general queries, individuals with self-stated ASD scanned more results than controls. The scrolled distance in the results page of general queries was 45% larger for the group of individuals with ASD (P<.001; d=0.45). The group of individuals with ASD also made the first scroll faster than the controls (P<.001; d=0.51). The differences in scrolled distance were larger for popular queries. No group differences in scrolled distance emerged for image queries, suggesting that visual load impeded the investigative behavior of individuals with ASD. No differences emerged in the number of clicks on search results.

Conclusions: Individuals who self-stated that they had ASD scrutinized more general search results and fewer image search results than the controls. Thus, our results at least partially support the notion that individuals with ASD exhibit investigative behaviors and suggest that textual searches are an important context for expressing such tendencies.

(J Med Internet Res 2021;23(7):e23829) doi:10.2196/23829

KEYWORDS

autism; decision making; exploration; search; internet

Introduction

Characterizing the internet browsing style of specific populations is important for understanding how individuals behave during naturally occurring circumstances and can ultimately be used for tailoring interfaces and content to people's diverse styles and capabilities [1,2]. This paper focuses on the internet search strategies of individuals with autism spectrum disorder (ASD)—a heterogeneous condition characterized by restricted interests and behaviors and by deficits in social communication.

XSL•FO
RenderX

Laboratory studies have discovered that high-functioning individuals with ASD tend to make selections from assorted options in a relatively short period of time [3-7]. For example, Johnson et al [3] examined the performance of adolescents and young adults with ASD in the Iowa gambling task [8]—a 4-choice, repeated, decision-making task. Individuals with ASD and typically developing controls had similar rates of selection with respect to the four options, but those with ASD switched options more frequently than controls. Similarly, in a simulated foraging task, Pellicano et al [5] found that children with ASD used more varied search paths across trials than those of control children. Herein, we analyze data from a large-scale data set and examine whether the heightened propensity for investigation in individuals with ASD emerges when they search for information on the web.

The internet has been described as a "newly autism-compatible environment" [9] due to fact that it is highly suitable for pursuing restricted interests [10] and because it allows individuals to control the mode of social communication [11]. Potentially, when searching for information in this environment, individuals with ASD may express their investigative tendencies by thoroughly examining search results. Alternatively, people with ASD tend to be more distracted by novel stimuli than typically developing persons [12,13]. They also tend to avoid situations involving multiple decisions [14,15]. If the novel stimulus and choice overload associated with ASD are sufficiently strong, they may restrict the examination of search results.

To discern whether individuals with self-stated ASD (s-sASD) examine internet search results more thoroughly than individuals without ASD, we analyzed their naturally occurring interactions with an internet search engine. Our research strategy was different from that of most autism studies, which typically use small samples of psychiatrically diagnosed participants [10,16,17]. We obtained historical browsing records for a very large group of users who stated that they had ASD in one of their search queries. We then used the same data set to filter identical queries made by users who did not refer to themselves as having ASD. Therefore, for a given query, we could compare the two groups' responses to the same query results and thus control for the content of the query (as well as any implied personality and cognitive style differences) and the specific search results. For robustness, we examined browsing records for two major domains—general web searches (study 1) and image searches (study 2).

Our main dependent variable was the scanned range of search results, that is, the extent to which individuals chose to present themselves with more results in addition to those shown on the initial results screen. In addition, we examined the number of clicks on search results (links to other pages) as an estimate of the exploitation of search outcomes. Finally, we tested whether there was a trade-off between increased search range and the time invested in viewing initial search results.

Methods

Study Design

This study was preapproved by the Technion Research Ethics Committee (approval number: 2020003). The data set used in this study is proprietary to Microsoft Corporation and was kept anonymous for privacy reasons. This data set was approved for research, which was to be conducted by the second author in his capacity as a Microsoft Research Senior Principal Researcher. By using a data set of all English-language queries made by people in the United States on the Bing search engine in 2019, we extracted each user's anonymized ID and their query text. Queries were filtered to identify users who described themselves as autistic in one of their queries. This was determined by the usage of one of the following phrases in query text: *I have autism*, *I'm autistic*, *I'm on the autism*, *I am on the autism*, and *my autism*. Queries that suggested that a user was unsure (eg, *do I have autism*) were excluded from the analysis. In addition, we excluded queries in which there was information suggesting that the reference to autism did not characterize the user (eg, *I have autism hat*). A complete list of exclusion terms appears in Multimedia Appendix 1. We referred to the group of users that we identified in this manner as individuals with s-sASD in 2019. The number of users in this group was 1746. The control group included all other users who were not identified as individuals with s-sASD and made queries on Bing in the United States in 2019.

Study 1: General Web Search

In study 1, we targeted all general web searches in Bing made by individuals with s-sASD in November 2019. Additionally, we examined identical searches that were made by the control group during the same time period. Our main dependent variable was the total distance scrolled (in pixels) in the search results screen, which denoted the extent to which additional results were scanned after the initially presented, on-screen results were scanned. For validation purposes, we also compared the number of scroll events. Scroll events are discrete increases in search range that have unspecified lengths and are produced by the user (using a keyboard or mouse). Additionally, we studied the number of clicked links in the search results. Furthermore, we compared the two groups' response times, that is, the amount of time until the first scroll event (ie, before the search range was increased) and, for control purposes, the time to the first mouse movement. We included unique queries that were searched at least 10 times. Additionally, we only included queries that were followed by mouse movements, which denoted a user's response. The total number of queries was 1491, and this represented over 300 million searches; 37,810 were conducted by individuals with s-sASD, and the rest were conducted by the control group.

Study 2: Image Search

In this study, we targeted all image searches in Bing made by individuals with s-sASD in November 2019 and all matching image searches made by the control group. Initially, in study 2a, we used the same search breadth variables as those in study 1—the total distance scrolled (in pixels), number of scroll events, and number of image clicks. As the criterion of

XSL•FO
RenderX

conducting 10 or more searches for a unique query produced a small amount of queries, we included unique image queries that were searched at least 5 times by both the study and control populations. Additionally, as previously stated, we only included queries that were followed by mouse movements. The total number of matching queries was only 38, which represented 364,386 searches (258 made by individuals with s-sASD). The relatively small number of matching image queries was likely due to the exclusion of devices such as tablets and cellphones, which do not have a mouse. In order to broaden the sample, in study 2b, we used an alternative dependent variable—the number of thumbnail images displayed to the user. This is similar to the total distance scrolled; however, it also includes the initial number of images that are available before scrolling. In this substudy, we also included searches with no mouse movements. The total number of image queries in study 2b was 78 (approximately 1.6 million searches, of which 698 were conducted by individuals with s-sASD).

Analysis

Search indices were averaged across each unique query. Each unique query therefore had a pair of data points—one data point set from the group of individuals with s-sASD and one from the control group. Query topic categories were identified by a proprietary classifier. Age and gender data were also available for a subset of users who were registered with Bing. Our main analysis involved a comparison of the differences between groups in terms of search indices across queries (using two-tailed paired *t* tests). This analysis was independent of the number of times that each query was issued. The analysis was therefore not confounded by the query topic because it examined responses that were contingent on a given query. In addition, to account for the effect of the popularity of searches, we also compared the average number of individual searches by multiplying the search parameters of each query by its relative volume (ie, the number of times that the query was issued divided by the average number of times a query was issued in a given group). If the difference between groups per query was not the same as the difference between individual searches, this implied that the volume of searches affected the gap between the two groups. In other words, this implied that the difference between groups was moderated by query popularity. The two differences were compared by using a repeated measures analysis of variance, in which the groups and units of analysis (queries vs searches) were paired factors.

Results

Search Queries

To further validate our self-statement method, we examined all Bing searches conducted by individuals with s-sASD in 2019 that included the words *autism* or *autistic*. On average, there were 14.17 such searches (SE 1.14). For comparison, we also examined those in the control group who made at least 1 search with the words *autism* or *autistic*. Their average number of relevant searches was only 2.52 (SE 0.003), and the difference in the average number of relevant searches between the two groups was highly significant (rank sum $P<.001$).

The top query topics of the individuals with s-sASD and the control group in November 2019 are shown in Table 1. Individuals with s-sASD were more likely to query about books and consumer electronics than controls and were more interested in images of automobiles and video games. In contrast, the control population made more queries related to travel and flights. We further examined whether the popularity of different queries was similar between the two groups (individuals with s-sASD and controls). The correlation between the volume of general queries included in this study for the two groups was 0.53 ($P<.001$), and for image searches, the correlation was 0.02 ($P=.86$). This implied that the relative search popularity of general queries was highly similar between groups, but this was not true for image-related queries.

XSL•FO
RenderX

Table 1. The top 3 query topics that were distinctly popular in either the group of individuals with s-sASD or the control group during November 2019. Odds ratios (ORs) and 95% CIs are presented.

Topics	OR (95% CI)
Topics that were more popular among individuals with self-stated autism spectrum disorder	
General queries	
Television shows	8.83 (8.49-9.18)
Books	6.16 (5.73-6.63)
Consumer electronics	4.36 (3.95-4.80)
Image queries	
Video games	4.20 (3.09-5.72)
Things to do	2.73 (2.42-3.08)
Automobiles	2.70 (2.46-2.95)
Topics that were more popular in the control group	
General queries	
Flights	4.82 (4.11-5.66)
Travel guide	4.70 (4.19-5.27)
Things to do	4.47 (4.04-4.94)
Image queries	
Travel guide	7.08 (5.52-9.10)
Restaurants	2.96 (2.50-3.52)
Television shows	1.69 (1.53-1.86)

Study 1: General Search

Table 2 shows the mean values of search parameters for the two groups, and Figure 1 presents the distribution of search parameters. Our main analysis of average values across different queries indicated that the scrolled distance was about 1.45 times higher among individuals with s-sASD (t_{1490}=8.63; P<.001; d=0.45), and the number of scroll events was about 1.12 times higher (t_{1490}=3.29; P=.001; d=0.17). In contrast, differences in the number of clicked links were smaller and not statistically significant (t_{1490}=1.91; P=.06; d=0.10).

Table 2. Study 1 results: the mean values of general search parameters for individuals with self-stated autism spectrum disorder (s-sASD) and the control groups.

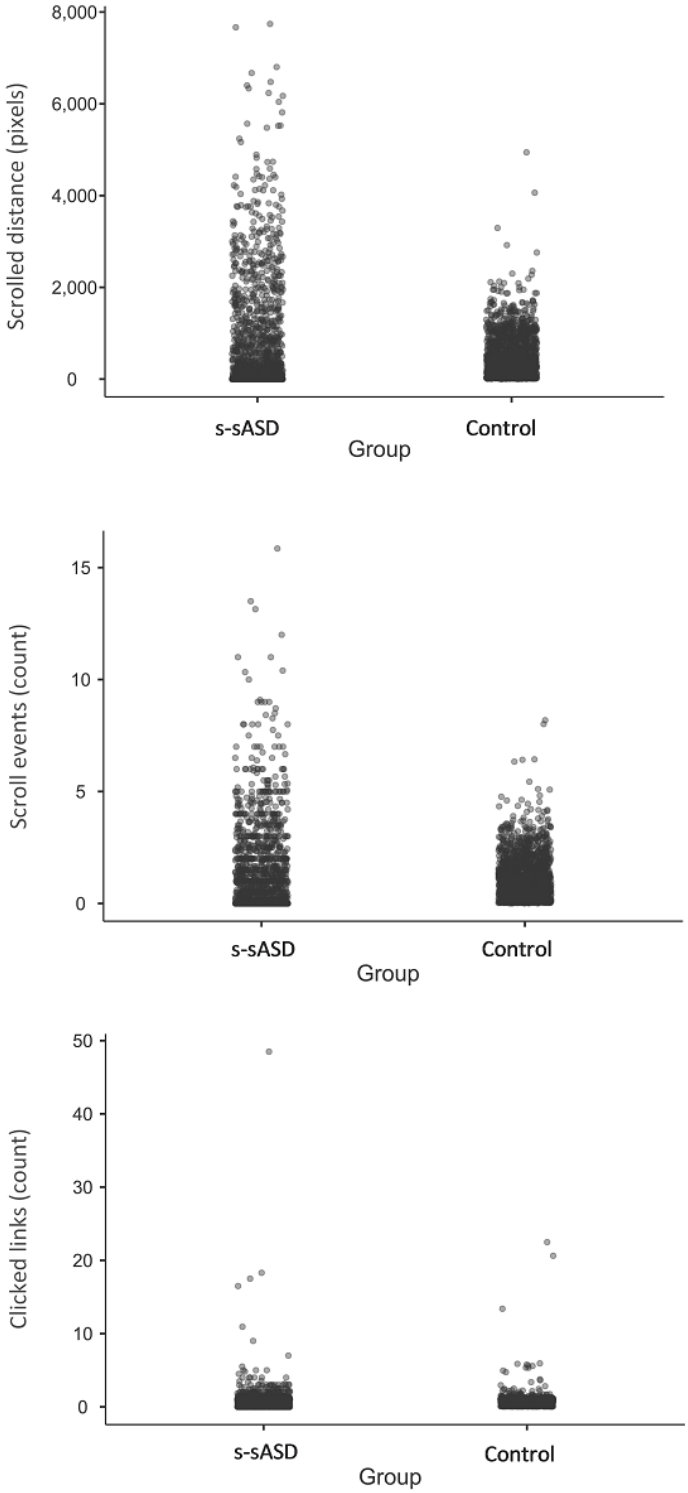
Variables	Values per query[a], mean (SE)		Average per individual search[b], mean (SE)		
	Individuals with s-sASD	Controls (matched)	Individuals with s-sASD	Controls (matched)	Controls (all other users)
Search breadth variables					
Scrolled distance (pixels)	745.38 (30.62)	513.44 (12.66)	835.29 (54.81)	136.07 (12.10)	231.16 (9.05)
Number of scroll events	1.33 (0.05)	1.18 (0.03)	1.35 (0.08)	0.32 (0.02)	0.55 (0.02)
Number of clicked links	0.94 (0.04)	0.86 (0.03)	1.11 (0.10)	0.93 (0.20)	0.86 (0.10)
Search time variables (s)					
Time to first scroll	35.99 (0.59)	40.81 (0.30)	33.83 (2.88)	51.34 (9.48)	48.74 (4.96)
Time to first mouse movement	3.28 (0.12)	3.91 (0.05)	3.21 (0.23)	4.48 (0.97)	4.14 (0.36)

[a]The average values for each query give the same weight to each query.

[b]The average values for each individual search were weighted by the relative volume of searches for each unique query.

XSL•FO
RenderX

Figure 1. Data for study 1 (search breadth indices stratified by group). The mean scrolled distance, number of scroll events, and number of clicked links for each query are presented. s-sASD: self-stated autism spectrum disorder.



The search breadth indices of the two groups were even more distinct when we compared individual searches (Table 2). The difference between the two groups' individual searches was significantly larger than the differences of each query in terms of scrolled distance ($F_{1,1490}$=98.36; P<.001) and the number of scroll events ($F_{1,1490}$=154.71; P<.001). However, this was not true for the number of clicked links ($F_{1,1490}$=0.36; P=.54). The scrolled distance of individual searches conducted by individuals with s-sASD was 6.14 times higher than that of the control group's searches (t_{1490}=13.91; P<.001; d=0.72), and the number of scroll events was about 4.28 times higher (t_{1490}=12.51; P<.001; d=0.71). This indicated that the queries' popularity moderated the differences in search breadth between groups, and the difference was higher for popular queries. To illustrate this effect, in Figure 2, we depict the differences in search indices among the top 50% of queries of both groups, those only in the control group, those only in the group of individuals with s-sASD, and those in neither group. As can be seen in Figure 2, a large difference between groups in terms of scrolled distance emerged in queries that were popular in both groups,

XSL•FO
RenderX

whereas a smaller difference emerged for queries that were relatively unpopular. Similarly, the difference between groups in terms of the number of scroll events mostly emerged for queries that were popular among individuals with s-sASD.

Figure 2. The effect of relative search volume on differences between groups in study 1. Queries were divided into the top 50% of queries in both groups, those in the control group, those in the group of individuals with s-sASD, and those in neither group. The error bars denote 95% CIs. s-sASD: self-stated autism spectrum disorder.

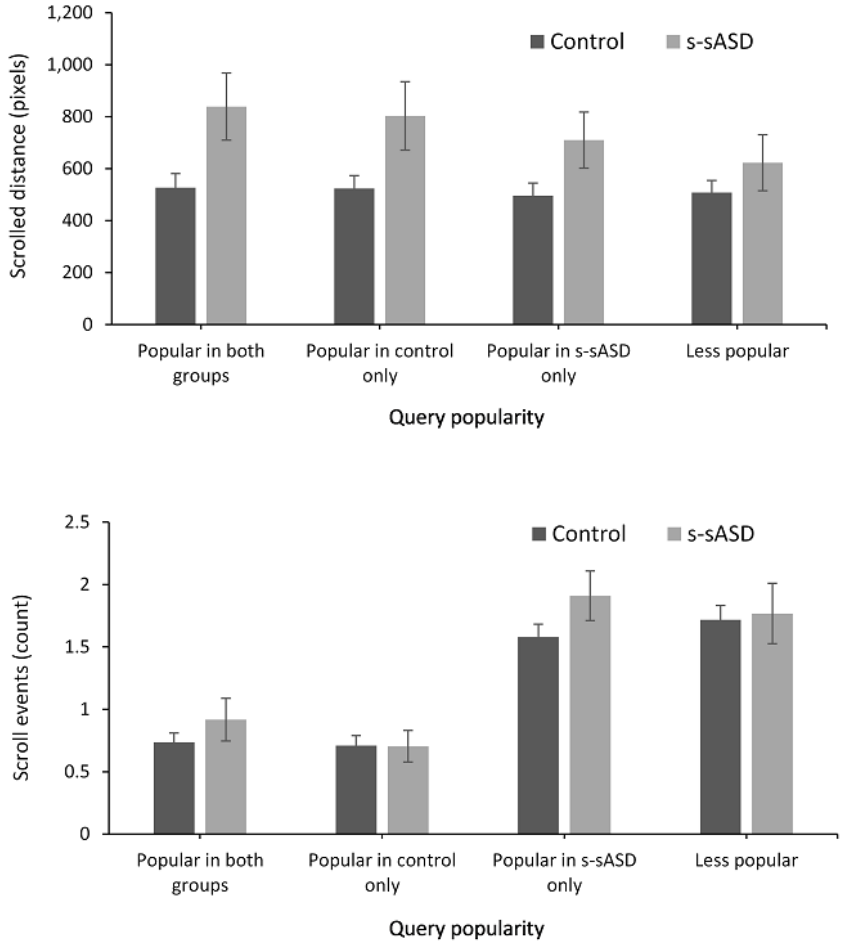


Table 2 also presents the average response times of the two groups. Individuals with s-sASD were considerably faster at making the first scroll event (t_{1490}=9.93; P<.001; d=0.51) and moving the mouse after conducting searches (t_{1490}=4.97; P<.001; d=0.26), suggesting that there was some trade-off between search breadth and the time invested in scanning the initial search results.

The differences between groups seemed consistent with the notion that ASD is associated with increased investigative behaviors. However, an alternative interpretation is that these differences were due to a demographic disparity between groups. Populations with ASD typically have a gender ratio of about 4:1 (males to females) [18]. Thus, it was important to determine if our results were moderated by gender. Since gender information was available only for a small subsample of users (less than 0.5% of the total users), we focused on participants in the control group and examined the effects that gender (0=female; 1=male) and age had on search indices by using linear regressions. A total of 5179 queries (approximately 7.0 million searches) were available for this analysis. The findings are summarized in Table 3. As can be seen in Table 3, gender had a rather small effect on the breadth of searches. Males' scrolled distance was higher than that of females by about 19 pixels, which equals 3.6% (pixels: 18.91/513.44) of the average scrolled distance of the control group. In contrast, the effect of s-sASD amounted to 45.2% (pixels: 231.94/513.44 pixels) of the average scrolled distance of the control group (pixels: 231.94/513.44). Thus, it appeared that the difference between genders was considerably smaller than the effect of self-stated autism.

Finally, because we matched the control group's queries to those of individuals with s-sASD, the studied queries represented 48.7% of the searches conducted by those with s-sASD and 15.2% of the searches conducted by the control group. In order to ensure that the subsample of searches conducted by the control group was not biased, we extracted all unique queries that were searched more than 10 times in the control group, except those that were included in our original sample (a total of 23,071 unique queries). We focused on indices for individual searches, since these queries were different from those included in this study. As can be seen in Table 2, the search breadth indices of all remaining queries in the control group were somewhat higher than those included in this study; however, they were still far below those recorded for the individuals with s-sASD.

XSL•FO
RenderX

Table 3. Study 1 results: an examination of the effects of age and gender in the control group. The adjusted r2 denotes the fit of the regression model (proportion of explained variance).

Predicted variables	Age, unstandardized coefficient (SE)	Gender, unstandardized coefficient (SE)	Adjusted r^2
Scrolled distance	0.07 (0.02)	18.91 (0.71)	0.004[a]
Scroll events	0.0008 (0.0001)	0.05 (0.002)	0.02[a]

[a]Significant at the $P<.001$ level.

Study 2: Image Search

Table 4 shows the mean values of image search parameters for the two groups, and the distribution of search parameters is presented in Figure 3. Different from study 1, in study 2a, we found that the scrolled distances were about equal between the two groups (t_{37}=0.43; P=.67; d=0.14). This was also true for the number of scroll events (t_{37}=0.32; P=.75; d=0.10). Additionally, on average, the group of individuals with s-sASD clicked on images considerably less than the control group (t_{37}=2.81; P=.008; d=0.91).

As in study 1, the tendency of participants with s-sASD to exhibit greater search breadth was more distinct when considering all individuals' searches rather than their average per query. This interaction was significant in terms of scrolled distance ($F_{1,37}$=9.05; P=.005) and the number of scroll events ($F_{1,37}$=5.58; P=.02). However, this was not true for the number of images clicked ($F_{1,37}$=4.05; P=.052). As indicated in Table 4, the mean scrolled distance in individual searches conducted by the individuals with s-sASD was higher than that of the control group (t_{37}=2.39; P=.02; d=0.78), as was the mean number of scroll events (t_{37}=2.36; P=.02; d=0.77). This was due to the fact that, as in study 1, for highly popular queries, individuals with s-sASD scanned more images than controls, and the number of individual searches for these queries was an order of magnitude higher than the median.

Similar findings emerged in study 2b, which included a somewhat larger number of queries (Table 4 and Multimedia Appendix 1). There was no significant difference in the number of images presented to the two groups per query (t_{77}=0.12; P=.90; d=−0.03). Additionally, in this large sample, the average number of image clicks in the group of individuals with s-sASD was smaller than that in the control group (t_{77}=4.43; $P<.001$; d=−1.00). The null effect of the number of images was similar when examining the average values for each query and average values for individual searches ($F_{1,77}$=2.81; P=.10).

We also examined whether the null effect in study 2b was due to queries mainly producing images of faces. We divided the image queries into those including individual persons' names (eg, "Bjork smiling": n=21; other queries: n=57) and compared the number of images presented in each category. The results indicated that for people-related searches, the control group was presented with, on average, 101.8 (SE 8.4) images, whereas 88.5 (SE 8.3) images were presented to the group of individuals with s-sASD. In contrast, for nonpeople-related searches, the control group was presented with 89.2 (SE 4.9) images, whereas 95.4 (SE 10.8) images were presented to the group of individuals with s-sASD. However, this crossover interaction trend was not significant in the repeated measures analysis ($F_{1,76}$=1.35; P=.25).

Table 4. Study 2 results: the mean values of image search parameters for individuals with self-stated autism spectrum disorder (s-sASD) and the controls.

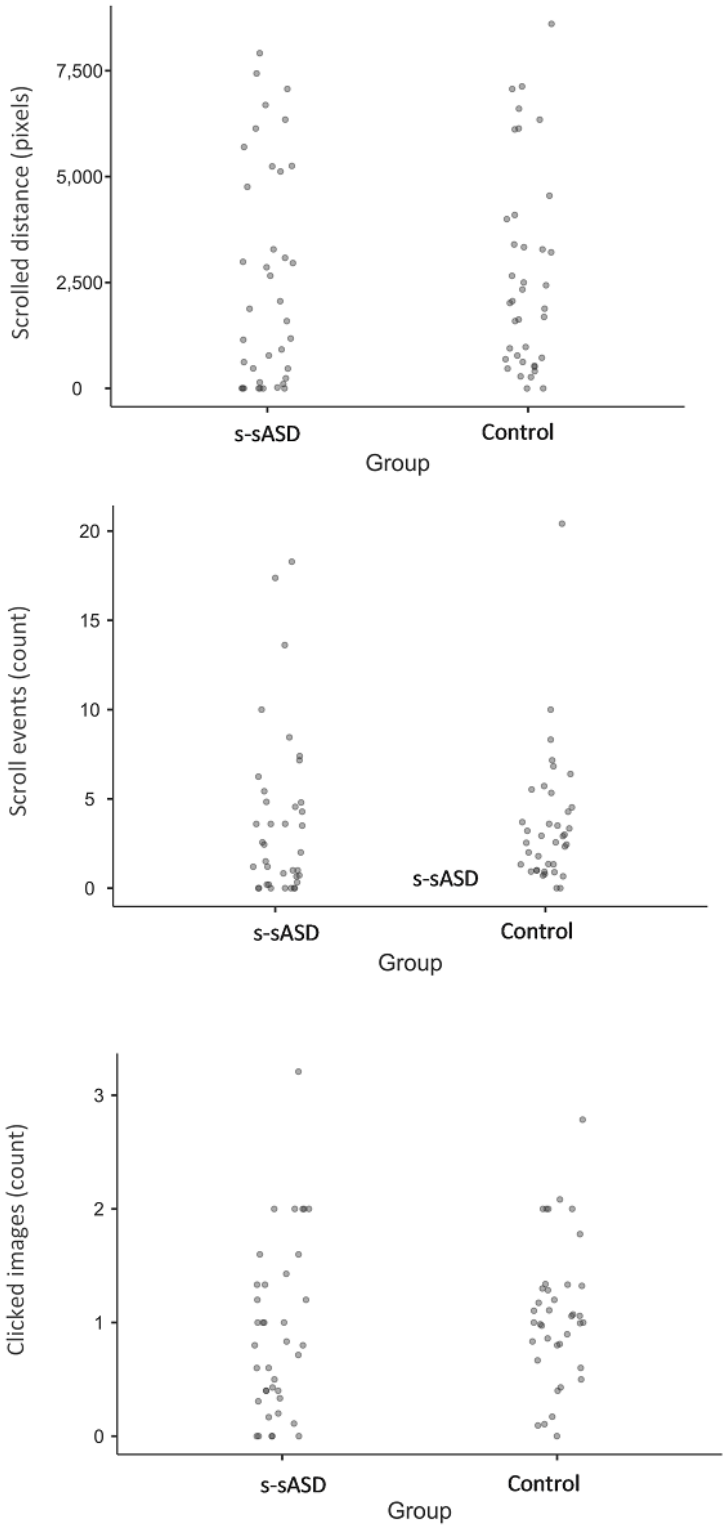
Variables	Values per query[a], mean (SE)		Values per individual search[b], mean (SE)	
	Individuals with s-sASD	Controls	Individuals with s-sASD	Controls
Study 2a variables				
Scrolled distance	2555.5 (419.5)	2680.5 (384.1)	3,111.7 (299.62)	668.23 (278.39)
Scroll events	3.75 (0.75)	3.56 (0.59)	4.43 (1.07)	1.03 (0.43)
Clicked images	0.90 (0.12)	1.08 (0.10)	1.09 (0.30)	0.28 (0.14)
Study 2b variables				
Images displayed	93.51 (8.16)	92.60 (4.24)	93.88 (13.32)	55.10 (17.39)
Clicked imaged	0.42 (0.06)	0.73 (0.07)	0.47 (0.11)	0.12 (0.04)

[a]The average values for each query give the same weight to each query.

[b]The average values for each individual search were weighted by the relative volume of searches for each unique query.

XSL•FO
RenderX

Figure 3. Data for study 2a stratified by group. The mean scrolled distance, number of scroll events, and number of clicked images for each query are presented. s-sASD: self-stated autism spectrum disorder.



Discussion

The first major finding we observed was that there was a large difference between how the individuals with s-sASD investigated search results for general searches and image searches. In general searches, participants with s-sASD scanned more results than controls by scrolling down on the screen of results. For example, for general queries, the scrolled distance of participants with s-sASD was 1.45 times larger than that of the control group (745.38 pixels vs 513.44 pixels). In image searches, no such trend emerged. The increased search range in general searches made by individuals with s-sASD supports the validity of laboratory studies showing that high-functioning individuals with ASD are prone to inspective behaviors [5,19]. However, the fact that this phenomenon was not found in such individuals conducting image searches suggests that there is an

XSL•FO
RenderX

important (and unexplored) moderator for the propensity to investigate autism.

The differences between image searches and general searches may have been due to the fact that image searches involve greater visual load than typical general searches. In individuals with ASD, the capacity for selective attention was found to be more impaired by high visual load when compared to that capacity in typically developing persons [20]. Additionally, image search results usually include more visual depictions of social interactions and faces, which individuals with ASD may find difficult to process [21-23]. Our findings showed that differences between groups were not significantly affected by whether the query was about an individual person, but additional individual-level analyses are necessary to further validate this finding.

The second major finding of this study was that the investigative tendency of individuals with s-sASD was more pronounced in individuals' average search parameters (across all of the queries they made) than in parameters averaged at the query level. This interaction was driven by an increase in the gap in highly popular queries between groups. For example, in study 1, with regard to the 20 most frequently searched queries in each group, the scrolled distance of individuals with s-sASD was 2.16 times higher than the mean scrolled distance of the control group. In contrast, with regard to the 20 least frequently performed queries, the scrolled distance for individuals with s-sASD was only about 1.26 times higher. Similar patterns emerged in study 2. The fact that differences between groups emerged mainly for queries that were highly popular among individuals with s-sASD is consistent with the principal tendency of such individuals to make large efforts for relatively specific areas of interest [24,25].

An important limitation of this study is the fact that individuals with s-sASD were not diagnosed with autism. Indeed, the only information about their identity stemmed from the statements that they made while searching for information on the web. Nevertheless, our validation analysis showed that these individuals were also actively making queries regarding autism-related websites. Furthermore, the favorite queries of individuals with s-sASD (Table 1) suggested that these individuals were more interested in technical topics than in social topics. Additionally, we controlled for the possibility that differences between groups might occur due to factors such as age and gender by examining the effect of these factors in the control group. Another limitation of this study is that it presumably represents a subsample of individuals with ASD who are capable of writing and are computer literate. Little is known about the search styles of low-functioning individuals with autism [26].

In our opinion, our findings make two contributions to the study of autism. First, they shed light on how individuals with self-reported ASD engage in naturally occurring internet searches. The answer appears to differ when the search mainly produces text and when it produces images, and these differences are affected by the queries' relative popularity. Therefore, an interesting open question is whether the web search style of individuals with autism, which involves the fast scanning of many search results, can be adapted by individuals with autism and those without autism.

Second, our findings provide a new method for assessing autism via the "digital footprints" of one's search statements. Therefore, we transitioned from the use of single or multiple keywords [27,28] to the use of self-referenced descriptors. Although we have relied on data that are not publicly available to achieve this goal, similar data can be extracted through crowdsourcing [29].

Conflicts of Interest

EY-T is an employee of Microsoft, owner of Bing.

Multimedia Appendix 1
List of exclusion terms.
[DOCX File , 12 KB - jmir_v23i7e23829_app1.docx]

References

1. Yom-Tov E, Brunstein-Klomek A, Mandel O, Hadas A, Fennig S. Inducing behavioral change in seekers of pro-anorexia content using internet advertisements: Randomized controlled trial. JMIR Ment Health 2018 Feb 22;5(1):e6 [FREE Full text] [doi: 10.2196/mental.8212] [Medline: 29472176]
2. Yoo DW, Birnbaum ML, Van Meter AR, Ali AF, Arenare E, Abowd GD, et al. Designing a clinician-facing tool for using insights from patients' social media activity: Iterative co-design approach. JMIR Ment Health 2020 Aug 12;7(8):e16969 [FREE Full text] [doi: 10.2196/16969] [Medline: 32784180]
3. Johnson SA, Yechiam E, Murphy RR, Queller S, Stout JC. Motivational processes and autonomic responsivity in Asperger's disorder: evidence from the Iowa Gambling Task. J Int Neuropsychol Soc 2006 Sep;12(5):668-676. [doi: 10.1017/S1355617706060802] [Medline: 16961948]
4. Yechiam E, Arshavsky O, Shamay-Tsoory SG, Yaniv S, Aharon J. Adapted to explore: reinforcement learning in Autistic Spectrum Conditions. Brain Cogn 2010 Mar;72(2):317-324. [doi: 10.1016/j.bandc.2009.10.005] [Medline: 19913345]

XSL•FO
RenderX

5. Pellicano E, Smith AD, Cristino F, Hood BM, Briscoe J, Gilchrist ID. Children with autism are neither systematic nor optimal foragers. Proc Natl Acad Sci U S A 2011 Jan 04;108(1):421-426 [FREE Full text] [doi: 10.1073/pnas.1014076108] [Medline: 21173235]
6. Mussey JL, Travers BG, Klinger LG, Klinger MR. Decision-making skills in ASD: performance on the Iowa Gambling Task. Autism Res 2015 Feb;8(1):105-114. [doi: 10.1002/aur.1429] [Medline: 25371315]
7. Vella L, Ring HA, Aitken MR, Watson PC, Presland A, Clare IC. Understanding self-reported difficulties in decision-making by people with autism spectrum disorders. Autism 2018 Jul;22(5):549-559 [FREE Full text] [doi: 10.1177/1362361316687988] [Medline: 28425298]
8. Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 1994;50(1-3):7-15. [doi: 10.1016/0010-0277(94)90018-3] [Medline: 8039375]
9. Murray D, Lesser M. Autism and Computing. 1999 Presented at: Autism '99. Online Conference Organized by the NAS with the Shirley Foundation; November 2, 1999; London, UK.
10. Jordan CJ, Caldwell-Harris CL. Understanding differences in neurotypical and autism spectrum special interests through Internet forums. Intellect Dev Disabil 2012 Oct;50(5):391-402. [doi: 10.1352/1934-9556-50.5.391] [Medline: 23025641]
11. Gillespie-Lynch K, Kapp SK, Shane-Simpson C, Smith DS, Hutman T. Intersections between the autism spectrum and the internet: perceived benefits and preferred functions of computer-mediated communication. Intellect Dev Disabil 2014 Dec;52(6):456-469. [doi: 10.1352/1934-9556-52.6.456] [Medline: 25409132]
12. Gomot M, Bernard FA, Davis MH, Belmonte MK, Ashwin C, Bullmore ET, et al. Change detection in children with autism: an auditory event-related fMRI study. Neuroimage 2006 Jan 15;29(2):475-484. [doi: 10.1016/j.neuroimage.2005.07.027] [Medline: 16115783]
13. Sokhadze E, Baruth J, Tasman A, Sears L, Mathai G, El-Baz A, et al. Event-related potential study of novelty processing abnormalities in autism. Appl Psychophysiol Biofeedback 2009 Mar;34(1):37-51. [doi: 10.1007/s10484-009-9074-5] [Medline: 19199028]
14. Luke L, Clare IC, Ring H, Redley M, Watson P. Decision-making difficulties experienced by adults with autism spectrum conditions. Autism 2012 Nov;16(6):612-621. [doi: 10.1177/1362361311415876] [Medline: 21846664]
15. Gaeth GJ, Levin IP, Jain G, Burke EV. Toward understanding everyday decision making by adults across the autism spectrum. Judgm Decis Mak 2016;11(6):537-546.
16. Daly BP, Nicholls EG, Patrick KE, Brinckman DD, Schultheis MT. Driving behaviors in adults with autism spectrum disorders. J Autism Dev Disord 2014 Dec;44(12):3119-3128. [doi: 10.1007/s10803-014-2166-y] [Medline: 24925544]
17. Levin IP, Gaeth GJ, Foley-Nicpon M, Yegorova V, Cederberg C, Yan H. Extending decision making competence to special populations: a pilot study of persons on the autism spectrum. Front Psychol 2015 Apr 28;6:539 [FREE Full text] [doi: 10.3389/fpsyg.2015.00539] [Medline: 25972831]
18. Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 2017 Jun;56(6):466-474. [doi: 10.1016/j.jaac.2017.03.013] [Medline: 28545751]
19. Zeif D, Yechiam E. Autism is not associated with poor or enhanced performance on the Iowa Gambling Task: A Meta-Analysis. Neurosci Biobehav Rev 2020 Jun;113:440-447. [doi: 10.1016/j.neubiorev.2020.04.016] [Medline: 32311370]
20. Tyndall I, Ragless L, O'Hora D. Effects of perceptual load and socially meaningful stimuli on crossmodal selective attention in Autism Spectrum Disorder and neurotypical samples. Conscious Cogn 2018 Apr;60:25-36. [doi: 10.1016/j.concog.2018.02.006] [Medline: 29522997]
21. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry 2002 Sep;59(9):809-816. [doi: 10.1001/archpsyc.59.9.809] [Medline: 12215080]
22. Norbury CF, Brock J, Cragg L, Einav S, Griffiths H, Nation K. Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. J Child Psychol Psychiatry 2009 Jul;50(7):834-842. [doi: 10.1111/j.1469-7610.2009.02073.x] [Medline: 19298477]
23. Vabalas A, Freeth M. Brief report: Patterns of eye movements in face to face conversation are associated with autistic traits: Evidence from a student sample. J Autism Dev Disord 2016 Jan;46(1):305-314. [doi: 10.1007/s10803-015-2546-y] [Medline: 26249261]
24. Gillberg IC, Gillberg C. Asperger syndrome--some epidemiological considerations: a research note. J Child Psychol Psychiatry 1989 Jul;30(4):631-638. [doi: 10.1111/j.1469-7610.1989.tb00275.x] [Medline: 2670981]
25. Ozonoff S. Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. Neuropsychology 1995;9(4):491-500. [doi: 10.1037/0894-4105.9.4.491]
26. Lambrechts A, Cook J, Ludvig EA, Alonso E, Anns S, Taylor M, et al. Reward devaluation in autistic children and adolescents with complex needs: A feasibility study. Autism Res 2020 Nov;13(11):1915-1928. [doi: 10.1002/aur.2388] [Medline: 32914568]
27. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]

XSL•FO
RenderX

28. Brigo F, Igwe SC, Ausserer H, Nardone R, Tezzon F, Bongiovanni LG, et al. Why do people Google epilepsy? An infodemiological study of online behavior for epilepsy-related search terms. Epilepsy Behav 2014 Feb;31:67-70. [doi: 10.1016/j.yebeh.2013.11.020] [Medline: 24361764]
29. Sadeh-Sharvit S, Fitzsimmons-Craft EE, Taylor CB, Yom-Tov E. Predicting eating disorders from internet activity. Int J Eat Disord 2020 Sep;53(9):1526-1533. [doi: 10.1002/eat.23338] [Medline: 32706444]

Abbreviations

ASD: autism spectrum disorder
s-sASD: self-stated autism spectrum disorder

Edited by R Kukafka; submitted 25.08.20; peer-reviewed by I Levin, WC Lee; comments to author 17.10.20; revised version received 26.10.20; accepted 24.05.21; published 06.07.21.

Please cite as:
Yechiam E, Yom-Tov E
Unique Internet Search Strategies of Individuals With Self-Stated Autism: Quantitative Analysis of Search Engine Users' Investigative Behaviors
J Med Internet Res 2021;23(7):e23829
URL: https://www.jmir.org/2021/7/e23829
doi:10.2196/23829
PMID:34255644

©Eldad Yechiam, Elad Yom-Tov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Characteristics of the Measurement Tools for Assessing Health Information–Seeking Behaviors in Nationally Representative Surveys: Systematic Review

Hanna Choi[1*], MSN, PhD; Gyeonghui Jeong[2,3*], MSc

[1]Department of Nursing Science, Nambu University, Gwangju, Republic of Korea
[2]College of Nursing, Chonnam National University, Gwangju, Republic of Korea
[3]College of Nursing, Seoul National University, Seoul, Republic of Korea
[*]all authors contributed equally

Corresponding Author:
Gyeonghui Jeong, MSc
College of Nursing
Chonnam National University
160 Baekseo-ro, Dong-gu
Gwangju, 61469
Republic of Korea
Phone: 82 10 4999 5110
Fax: 82 62 227 4009
Email: gyeonghui.jeong@gmail.com

Abstract

Background: The coronavirus pandemic (COVID-19) has also emerged as an infodemic, thereby worsening the harm of the pandemic. This situation has highlighted the need for a deeply rooted understanding of the health information–seeking behaviors (HISBs) of people.

Objective: The aim of this paper was to review and provide insight regarding methodologies and the construct of content in HISB surveys by answering the following research question: what are the characteristics of the measurement tools for assessing HISBs in nationally representative surveys around the world?

Methods: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses was used as the framework for this study. A data search was performed through 5 international and 2 Korean databases covering the years between 2008 and 2020. Initially, studies performed among nationally representative samples were included to discover HISB survey instruments. The methodologies of the studies using HISB surveys were analyzed. For content analysis, 2 researchers reached a consensus through discussion by scrutinizing the contents of each survey questionnaire.

Results: A total of 13 survey tools from 8 countries were identified after a review of 2333 records from the search results. Five survey tools (Health Information National Trends Survey, Health Tracking Survey, Annenberg National Health Communication Survey, National Health Interview Survey, and Health Tracking Household Survey) from the United States, 2 instruments from Germany, and 1 tool from each of the countries of the European Union, France, Israel, Poland, South Korea, and Taiwan were identified. Telephone or web-based surveys were commonly used targeting the adult population (≥15 years of age). From the content analysis, the domains of the survey items were categorized as follows: information (information about health and patient medical records), channel (offline and online), and health (overall health, lifestyle, and cancer). All categories encompassed behavioral and attitude dimensions. A theoretical framework, that is, an information-channel-health structure for HISBs was proposed.

Conclusions: The results of our study can contribute to the development and implementation of the survey tools for HISB with integrated questionnaire items. This will help in understanding HISB trends in national health care.

(J Med Internet Res 2021;23(7):e27539) doi:10.2196/27539

XSL•FO
RenderX

KEYWORDS

information seeking behavior; consumer health information; medical informatics; health care surveys; health information–seeking behavior; surveys

Introduction

Background

The recent global pandemic of COVID-19, determined to be a public health emergency of international concern, has changed many aspects of people's daily lives [1]. When people wake up, they check health-related news, their signs and symptoms, methods of prevention, and restrictions on the use of a vaccine. While mass media have been releasing a myriad of information, individuals have also been reproducing and downloading news and information from internet webpages such as websites or blogs [2,3]. The tsunami of information has resulted in the production of several fake news that lack scientific evidence and convey misconceptions and misinformation about health [4]. In reality, misguided belief based on misinformation has caused the deaths of many people [5] and worsened COVID-19 infections [6,7]. In this way, the rise of incorrect information has led to abuse, or in other words, an infodemic [4,8]. The foremost solution to mitigate this issue would be to understand the information-seeking behaviors of individuals. It would be beneficial if governments or national institutes measure their behaviors to apply health and information policies appropriately [9].

Health information–seeking behavior (HISB) is a comprehensive term that describes an individual's behavior of seeking information, including the intentional collection and unintentional receipt of information [10,11]. Some studies have shown HISBs by using certain measurement tools such as Health Information National Trends Survey (HINTS), Health Tracking Survey, and the Annenberg National Health Communication Survey (ANHCS). The limitations of these studies are that most surveys mainly target American subjects or web-based/digital HISB [12-18]. These limitations can be overcome by the design of a comprehensive survey instrument. Survey instruments are developed to collect information for certain research phenomena [19] or for finding the right answers by asking the right questions. It would be efficient and effective to obtain a holistic view by integrating the properties of worldwide national survey tools in a systematic approach and by scrutinizing the constructs and methodologies, including what aspects of HISBs are considered important or are missed out. Although there are preliminary studies using systematic reviews of HISB instruments, these topics are limited to the context of the United States and eHealth, thereby making it difficult to look into cross-national HISB [17,20]. Therefore, this study aims to review how HISBs are measured by identifying and comparing measurement tools based on nationwide surveys.

Objectives

The aim of this paper was to provide insights on the methodologies and the construct of content for HISB survey instruments based on nationally representative surveys.

Methods

Research Question

The SPIDER (sample, phenomenon of interest, design, evaluation, and research type) format was used to formulate the research question for this review [21,22]: what are the characteristics of measurement tools (evaluation) for assessing HISBs (phenomenon of interest) in nationally representative surveys around the world (sample and design)?

Protocol and Registration

This study was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses [23]. The protocol of this review paper is registered in PROSPERO (CRD42019122767).

Eligibility Criteria

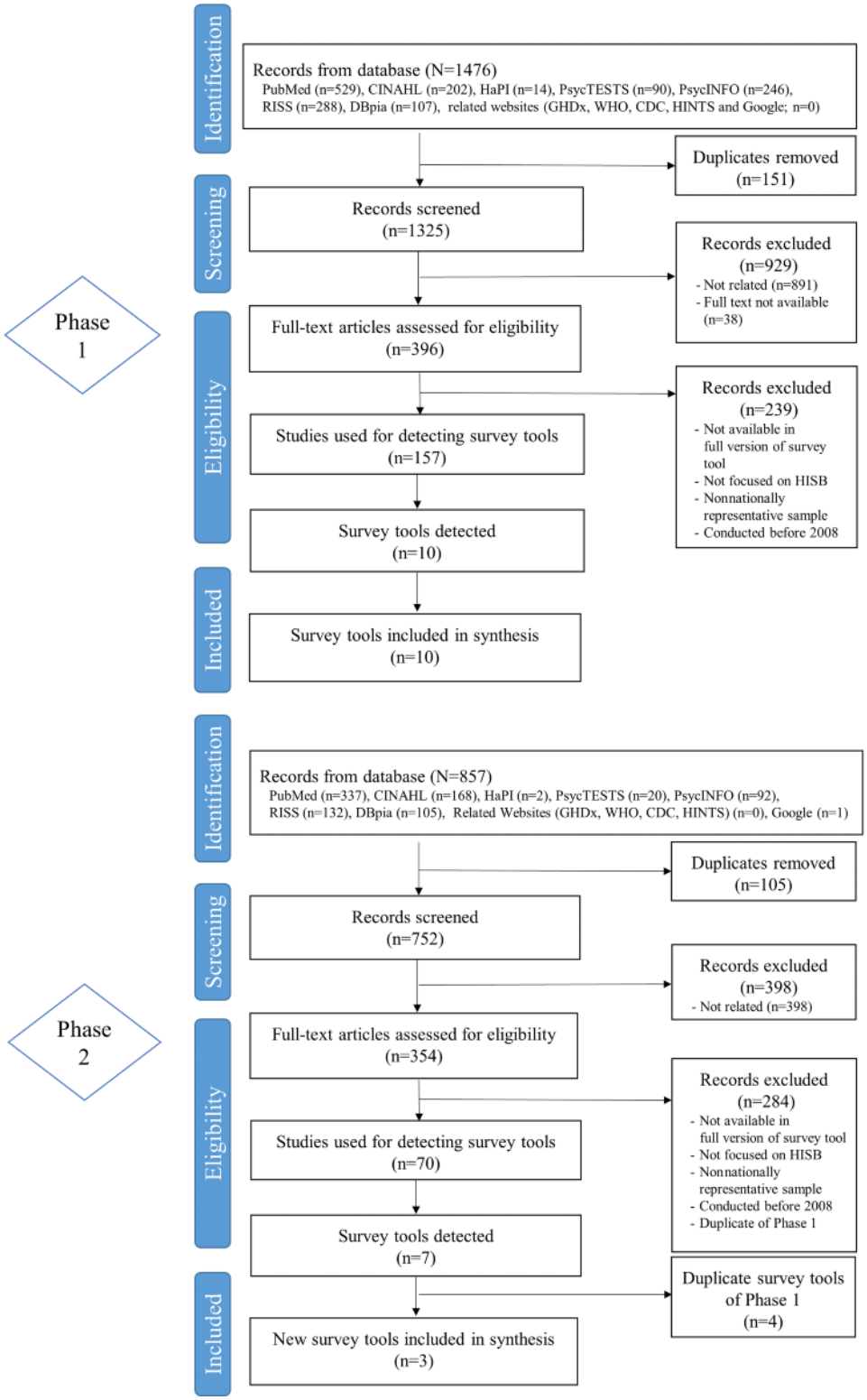
To answer the research question, inclusion and exclusion criteria were established. Survey tools were included if they were full versions of the tools for HISBs and if they targeted nationally representative samples. However, tools were excluded when the full versions of the instruments were not accessible, not HISB-focused, nor used for a nationally representative sample.

Information Sources

As we seek in this study to discover the national survey tools for HISB, articles, reports, and related websites were searched for clues to detect those instruments. The data search was performed in 2 phases. The phase 1 search covering 2008 to 2017 was conducted between October 09, 2017 and November 13, 2017 through 7 databases: 5 international databases, namely, PubMed, CINAHL Complete (Ebsco), HaPI, PsycTESTS, and PsycINFO (Ebsco), and 2 Korean databases (RISS [Research Information Sharing Service] and DBpia). Phase 2 was performed between February 19, 2021 and March 25, 2021 to obtain recent literature covering 2017 to 2020 with the same search strategy (Figure 1, Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. PRISMA flow diagram of literature search and selection process. CDC: Centers for Disease Control and Prevention; HINTS: Health Information National Trends Survey; HISB: health information–seeking behavior; RISS: Research Information Sharing Service; WHO: World Health Organization.



Search Strategy

Pilot searches were performed by the authors, and the final search strategy with the consultation of a librarian was utilized with MeSH terms (ie, information-seeking behavior) and free-text searching as well as the Boolean operators “OR” and “AND” (Multimedia Appendix 1). There was no limit on languages, but publication years were restricted between 2008 and 2020: January 1, 2008 to November 13, 2017 for phase 1 and January 1, 2017 to December 31, 2020 for phase 2.

Study Selection and Data Collection Process

Two authors (HC and GJ) initially reviewed the titles and abstracts of the papers and eliminated irrelevant documents. Then, HC and GJ scrutinized full-texts and filtered them according to the inclusion/exclusion criteria. As the purpose of

XSL•FO
RenderX

the study was to seek nationally representative surveys of HISB, related websites were also accessed, such as that of The World Health Organization, which has the primary role of directing and coordinating international health, and Global Health Data Exchange [24], which is the most comprehensive catalog of surveys, censuses, vital statistics, and other health-related data in the world. In addition, to obtain the survey questionnaires, websites such as those of the National Cancer Institute, Centers for Disease Control and Prevention, European Commission, and Santé Publique France were searched. Academic papers, reports, and webpages identified through the previous steps were reviewed to discover HISB survey tools. To attain sufficient data (ie, full version of the item(s) of the survey, methodology, etc), we emailed 8 corresponding authors of the papers: 2 of the corresponding authors sent full version of the survey instruments, which were not related to the HISB; 1 author refused to provide a full version of the survey instrument; and 5 authors did not respond. To capture grey literature, footnote tracing was performed along with a review of the related websites described above. All documents identified through this process were managed with EndNote X20.0 software (Clarivate Analytics). During the whole process, consensus was reached through discussion if there was disagreement between the authors.

Data Items

We sought the characteristics of the selected instruments, including the name of the instrument, administrative institution, and funding sources, country, language, frequency of the survey, survey duration, sampling method, mode of survey administration, target population, total number of the population, and purpose of the measurement. In addition, the content of the survey instruments was scrutinized.

Risk of Bias in Individual Studies

The aim of this study was to identify the measures used to analyze HISB in national surveys. Therefore, this review paper focuses on questionnaires in the national surveys on HISB and the risk of bias assessment is not applicable.

Synthesis of Results

As this review is intended as content analysis, the authors thoroughly read the contents of the questionnaires of the selected HISB instruments. Themes emerged during this process as we used coding sheets with Excel and Word. The findings were provided through the process of reaching a consensus between the 2 authors on the coding sheets. Finally, the synthesized results were depicted in table and figure formats.

Results

Study Selection

A total of 2333 papers were identified through 2 phases of the search process. From phase 1 of the search, 1476 papers were identified in the following academic databases: PubMed (n=529), CINAHL (n=202), HaPI (n=14), PsycTESTS (n=90), PsycINFO (n=246), RISS (n=288), and DBpia (n=107). Duplicates (n=151) were removed and 929 papers were eliminated. A total of 396 full-texts were reviewed and 157 documents were used for detecting 10 survey tools: (1) HINTS [25], (2) Health Tracking Survey [26], (3) ANHCS [27] (n=5), (4) National Health Interview Survey (NHIS) [28], (5) Health Tracking Household Survey (HTHS) [29], (6) Flash Eurobarometer [30], (7) Baromètre Santé [31], (8) Gesundheitsmonitor [32], (9) Israeli survey [33], and (10) eHealth Consumer Trend Survey [34].

Phase 2 was performed to update the recent survey tools by using the same search strategy. As a result, 857 records were identified: PubMed (n=337), CINAHL (n=168), HaPI (n=2), PsycTESTS (n=20), PsycINFO (n=92), RISS (n=132), DBpia (n=105), and Google (n=1). Duplicates (n=105) were excluded, and 398 records were also removed after screening. The full texts of 354 papers were reviewed, and 70 records were used for detecting 7 survey tools. There were 4 duplicates of survey tools from phase 1. Therefore, 3 more survey tools, that is, Stiftung Gesundheitswissen (HINTS Germany) [35], survey of cancer and health-related information–seeking behavior (CHISB) for Koreans [36], and Taiwan Communication Survey [37] were also included for synthesis.

A total of 227 papers were related to the selected HISB instruments (Multimedia Appendix 2). About 96% of them (219/227) were related to 1 of the 5 US surveys: HINTS [25] (n=188), the Health Tracking Survey [26] (n=9), ANHCS [27] (n=7), NHIS [28] (n=11), and HTHS [29] (n=4). The remaining 8 studies identified 8 survey tools used in other parts of the world, that is, European Union (Flash Eurobarometer) [30] (n=1), France (Baromètre santé) [31] (n=1), Germany (Gesundheitsmonitor [32] [n=1] and HINTS Germany) [35] [n=1]), Israeli survey [33] (n=1), Poland (eHealth Consumer Trend Survey) [34] (n=1), South Korea (survey of CHISB) [36] (n=1), and Taiwan (Taiwan Communication Survey) [37] (n=1). Therefore, 13 survey instruments (Table 1, Multimedia Appendix 3) were included in this review [38-77].

XSL•FO
RenderX

Table 1. Brief characteristics of the instruments for measuring health information–seeking behaviors in nationally representative survey studies.

Country	Instrument	Survey version	Purpose of the measurement	Frequency	Target population	Total population in the survey (N)
USA	Health Information National Trends Survey (HINTS) [38-41]	2019, HINTS 5, Cycle 3	To investigate respondents' access to and use of health information, including information technology to manage health and health information	Every few years (1-2 year cycle)	Civilian noninstitutionalized adults aged 18 years or older	5247
USA	Health Tracking Survey [42-50]	2012	To assess pursuit of health taking place within a widening network of both online and offline sources	Irregular	Adults aged 18 years or older	3014
USA	Annenberg National Health Communication Survey [18,51-56]	2012	To capture national trends related to health behavior and behavioral intentions to media exposure, health knowledge and beliefs, and policy preferences and beliefs	One-cycle survey	Adults aged 18 years or older	3692
USA	National Health Interview Survey [57-67]	2020	To monitor the health of the population through the collection and analysis of the data	Annual	Household	33,138[a]
USA	Health Tracking Household Survey [68-71]	2010	To inform health care decision makers about changes in the health care system and the influence	Irregular (2-5 year period)	Household	16,671 individuals (n=9165 Family Insurance Units)
Europe	Flash Eurobarometer 404 (European citizen's digital health literacy) [72]	2014	To support increasing use of digital health care to help manage citizen's own health	One-cycle survey	EU residents aged 15 years and older	26,566 (28 EU countries)
France	French Health Barometer (Baromètre santé) [73]	2017	To gain a better understanding of French health knowledge, attitudes, beliefs, and behaviors	Annual	Adults aged 18-75 years	15,635[b]
Germany	Gesundheitsmonitor [74]	2015	To assess health-related knowledge, attitudes, and behaviors	Annual	Adults aged 18-79 years	1598
Germany	HINTS Germany [75]	2019	To close the gap in important health-related information actions and systematical health records	Every few years (1-2 year cycle)	Adults aged 18-79 years	2902
Israel	Not titled survey [33]	2014	To measure eHealth literacy for others, including perceived outcome of internet use	One-cycle survey	Adult aged 21 years and older	819
Poland	eHealth Consumer Trend Survey 2012[c] [76]	2012	To show the trends in the perceptions and preferences of Polish citizens regarding internet use and factors affecting their usage	Irregular	Adults aged 15-80+ years	1000
South Korea	Survey of cancer and health-related information–seeking behavior for Koreans [36]	2018	To capture national phenomena of cancer and health-related health information–seeking behavior of Koreans	One-cycle survey	Adults aged 18-65+ years	1012
Taiwan	Taiwan Communication Survey [77]	2016	To explore media use behaviors among the general public, including health, risk, and disaster communication	Annual	Adults aged 18 years and older	2098

[a]2019 sample size was reported. Data and report for 2020 will be published in fall 2021.

[b]French Health Barometer: the survey questionnaires were changed according to the survey years. The 2017 version of the survey contains health information–seeking behavior and is included in this study.

[c]eHealth consumer trend survey of 2012 was modified from the eHealth Consumer Trends Survey (2007), which was conducted in Denmark, Germany, Greece, Latvia, Norway, Poland, and Portugal in the World Health Organization/European eHealth Consumer Trends project [78,79].

XSL•FO
RenderX

Key Characteristics of the Surveys

Country

HISB surveys were found in 8 countries (Table 1, Multimedia Appendix 3). The United States has 5 HISB surveys (HINTS, Health Tracking Survey, ANHCS, NHIS, and HTHS), and the other 7 countries or regions, namely, the European Union, France, Germany, Israel, Poland, South Korea, and Taiwan conduct surveys called Flash Eurobarometer, Baromètre santé, Gesundheitsmonitor, Israeli survey (not titled), the eHealth consumer trend survey, survey of CHISB for Koreans, and Taiwan Communication Survey, respectively.

Language

As the surveys focused on domestic people, official or national languages were used (Table 1, Multimedia Appendix 3). For instance, the HINTS from the United States used 2 versions of the survey: English and Spanish. The European Union also performed the survey using the mother tongue of the responders.

Instrument and Administration Institution

HISB surveys were administered by national, nonprofit, public institutions, or individual researchers (Table 1, Multimedia Appendix 3). Five instruments, that is, HINTS, Flash Eurobarometer, NHIS, Baromètre santé, and Taiwan Communication Survey, were developed and administered by national institutes, namely, the National Cancer Institute in the United States, the National Center for Health Statistics in the United States, the Directorate-General for Communications Networks of the European Commission, the National Institute for Prevention and Health Education in France, and the Ministry of Science Technology in Taiwan, respectively. Four instruments were obtained from nonprofit institutions: the Pew Research Center (HINTS), the Center for Studying Health System Change (ceased operation in 2013) (HTHS), Bertelsmann Stiftung (Gesundheitsmonitor), and Gesundheitswissen and Hanover Center for Health Communication at the Institute for Journalism and Communication Research (HINTS Germany). A survey (ANHCS) was conducted by 2 public institutions, namely, the Annenberg Schools for Communication at the University of Pennsylvania and the University of Southern California. Individual researchers developed 3 survey tools: the Israeli survey, the eHealth Consumer Trend Survey (Poland), and the survey of CHISB for Koreans (South Korea), with the Israeli and South Korean studies funded by national institutes.

Frequency of the Survey

The frequency of the surveys was found to be annual, every few years, one time, or irregular (Table 1, Multimedia Appendix 3). The annual or every few years surveys were HINTS (United States), NHIS (United States), Baromètre santé (France), Gesundheitsmonitor (Germany), HINTS Germany (Germany), and Taiwan Communication Survey (Taiwan). The others, namely, the Health Tracking Survey (United States), ANHCS (United States), HTHS (United States), Flash Eurobarometer 404, the Israeli survey, survey of CHISB for Koreans (South Korea), and eHealth Consumer Trend Survey (Poland) have been conducted once or irregularly.

Sampling and Mode of Administration

The most common approach has been randomization, in particular, sampling with random digit dialing and then administration through a computer-assisted telephone interview (Table 1, Multimedia Appendix 3). In addition, for sampling, two-stage sampling (stratifying sample addresses and selecting 1 adult within each household) was often used. When web-based panels were used for random sampling, units or strata layers divided by the population group, geographical districts, size of the settlement, and the locality's socioeconomic status were utilized to prevent clashes.

Population

The range of this study is restricted to researching tools used with adults (Table 1, Multimedia Appendix 3). The standard age of adulthood in each country varies from 15 years to 21 years. Mostly, adults are defined as people who are 18 years of age or older, but in Europe and Poland, those who are 15 years or older are considered part of the adult population. In Israel, people older than 21 years are considered adults.

Purpose

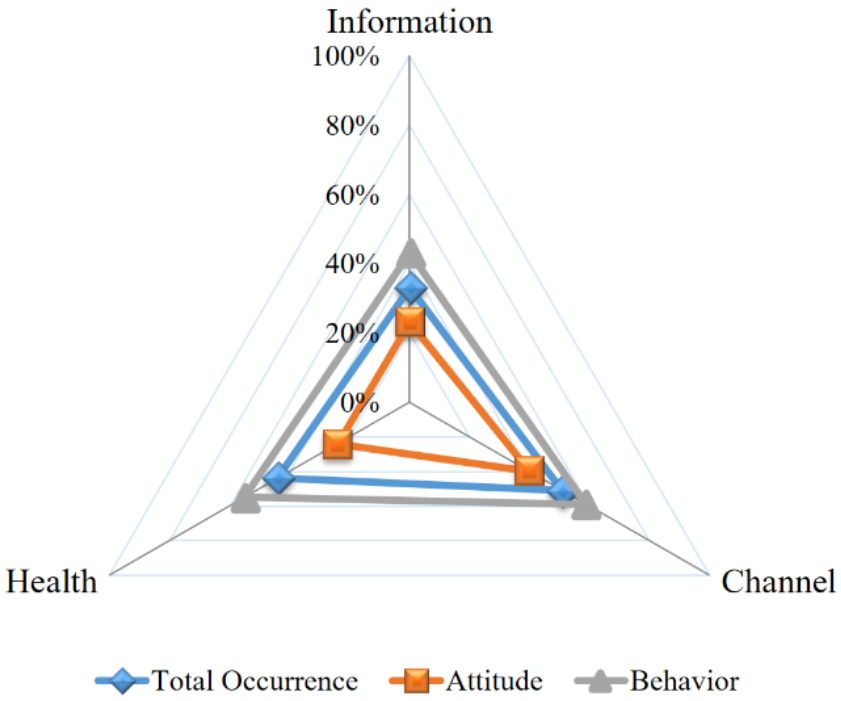
The purposes were similar among the measurements: to monitor the use of health information in accordance with the type of information technology such as online or offline (Table 1, Multimedia Appendix 3). However, the detailed outcome of the studies pursued was different. For instance, the Baromètre santé (France) aimed to discover knowledge, attitudes, and behaviors toward HISB; however, the ANHCS (United States) pursued HISB related to media exposure, health knowledge and beliefs, and policy preferences and beliefs.

Content Analysis of the Instruments

The contents of the questionnaire items for each tool were thematically reviewed and categorized by 2 researchers (HC and GJ). The themes were then merged and synthesized through consensus. Thus, 57 themes were detected and divided into 3 domains (Figure 2) and 7 subdomains: information, information about health and patient medical records; channel, offline and online; and health, overall health, lifestyle, and cancer. Two dimensions—attitude and behavior—were identified across the domains (Table 2, Multimedia Appendix 4). In this paper, attitude was defined as the emotional and cognitive tendency of a person toward a particular object, person, or thing, affecting behavior [80]. Behavior was also defined as an objectively observable activity [81].

XSL•FO
RenderX

Figure 2. Average percentage of theme occurrence in the domains.



XSL•FO
RenderX

Table 2. Content analysis of 13 representative national health information–seeking tools.[a]

Domain, subdomain, dimension, theme	Theme occurrence (%)	Theme occurrence average percentage (SD)	Subdomain average percentage (SD)	Domain average percentage (SD)
Information				33.0 (14.9)
Information about health			44.9 (14.9)	
Attitude		26.9 (5.4)		
Perceived ease of use	30.8			
Perceived efficacy of seeking	23.1			
Behavior		53.8 (6.3)		
Search experience (frequency)	46.2			
Information source	61.5			
Type of information contents	53.8			
Purpose of search (for whom)	53.8			
Patient medical record			24.0 (6.4)	
Attitude		21.5 (6.4)		
Perceived privacy and confidentiality risk	23.1			
Perceived ease of use	15.4			
Perceived usefulness	23.1			
Intention to use	15.4			
Preference to provide access to others	30.8			
Behavior		28.2 (4.4)		
Access frequency	30.8			
Type of information contents sought	23.1			
Purpose of seeking a record	30.8			
Channel				50.5 (18.2)
Offline			50.5 (15.9)	
Attitude		41.0 (4.4)		
Perceived credibility	38.5			
Perceived ease of use	38.5			
Satisfaction with service quality	46.2			
Behavior		57.7 (18.3)		
Access frequency	84.6			
Type of health service	46.2			
Communication with health care provider	46.2			
Health-related decision making	53.8			
Online			50.5 (19.7)	
Attitude		39.6 (15.0)		
Perceived credibility	53.8			
Perceived ease of use	38.5			
Perceived usefulness	53.8			
Perceived eHealth literacy (technology efficacy)	53.8			
Satisfaction with web-based information	15.4			
Perceived confidentiality risks	30.8			
Intention to use	30.8			

XSL•FO
RenderX

Domain, subdomain, dimension, theme	Theme occurrence (%)	Theme occurrence average percentage (SD)	Subdomain average percentage (SD)	Domain average percentage (SD)
Behavior		59.0 (19.2)		
Access frequency	92.3			
Type of information technology device	61.5			
Health-related web and app (software use)	46.2			
Web-based resource (governmental website, Wikipedia, etc)	53.8			
Communication (consult) with health care provider	76.9			
Communication with friends and others (social media, forum, etc)	61.5			
Health-related decision making	69.2			
Tracking/managing health state	38.5			
Improvement of health knowledge	30.8			
Health				44.2 (20.6)
Overall health			53.8 (18.0)	
Attitude		34.6 (5.4)		
Perceived health efficacy	38.5			
Concerns and belief about health	30.8			
Behavior		59.3 (16.4)		
General health state	84.6			
Diseases diagnosed	69.2			
Height	61.5			
Weight	61.5			
Mental health	53.8			
Caregiving	30.8			
Social support	53.8			
Lifestyle			32.7 (21.3)	
Attitude		15.4 (8.9)		
Perception about nutrition	7.7			
Perception about physical activity	15.4			
Perception about alcohol	23.1			
Perception about tobacco	23.1			
Behavior		50.0 (13.3)		
Nutrition	30.8			
Physical activity	61.5			
Alcohol	53.8			
Tobacco	53.8			
Cancer			46.2 (13.3)	
Attitude				
Perception about cancer	38.5			
Behavior		50.0 (16.3)		
Cancer check-up	38.5			

XSL•FO
RenderX

Domain, subdomain, dimension, theme	Theme occurrence (%)	Theme occurrence average percentage (SD)	Subdomain average percentage (SD)	Domain average percentage (SD)
Cancer diagnosed	61.5			

[a]Total average percentage of the themes=44.0 (SD 19.3), total average percentage of attitude themes=30.4 (SD 13.5), and total average percentage of behavior themes=53.8 (SD 16.9).

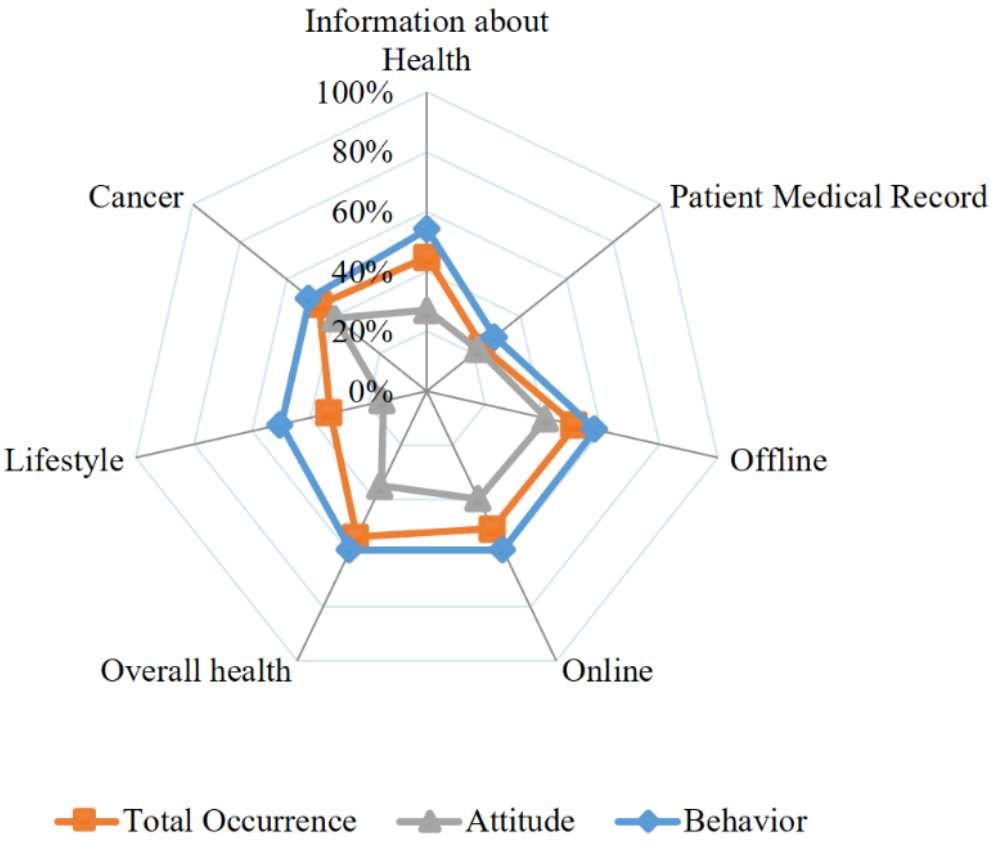
Thematic Map

Three domains, namely, information, channel, and health (Figure 2) emerged through the content analysis (Table 2). The highest rate of theme occurrence among the domains was channel (average percentage 50.5%, SD 18.2), followed by health (average percentage 44.2%, SD 20.6) and information (average percentage 33.0%, SD 14.9).

Information

Information is a health-related, content-focused domain sought by the individual. There are 2 subdomains (Figure 3), namely, information about health and patient medical records. The information about the health subdomain was conceptualized by categorizing question items related to general health information through a set of options with comprehensive channels (online or offline). Patient medical records were related to a seeker's use of medical records online or offline. There were attitude and behavioral aspects for the themes found, and the detailed and representative questionnaire items of the themes are presented in Table 3. The subdomain information about health (average percentage 44.9%, SD 14.9), which consisted of 6 themes, was more commonly used among the selected tools than patient medical records (average percentage 24.0%, SD 6.4), which consisted of 8 themes. In both subdomains, the percentages of behavior-related themes was 1.3-2.0 times higher than those related to attitude.

Figure 3. Average percentage of theme occurrence in the subdomains.



XSL•FO
RenderX

Table 3. Representative sample questionnaire items for health information–seeking behavior survey instruments.

Domain, subdomain, dimension, theme	Questionnaire items
Information	
Information about health	
Attitude	
Perceived ease of use	How much do you agree or disagree- it took a lot of effort to get the information you needed (HINTS[a])
Perceived efficacy of seeking	How confident are you that you could get advice about health if you needed it (HINTS)
Behavior	
Seek experience	Have you ever looked for information about health or medical topics from any source? (HINTS)
Information source	Thinking about the last time you had a serious health issue, did you get information from (selection of the information source)? (HTS[b])
Type of information contents	What type of health-related information did you look for? (Europe)
Purpose of search (whom for)	The most recent time you looked for information about health or medical topics, who was it for? (HINTS)
Patient medical record	
Attitude	
Perceived privacy and confidentiality risk	Have you ever kept information from your health care provider because you were concerned about the privacy or security of your medical record? (HINTS)
Perceived ease of use	How easy or difficult was it to understand the health information in your online medical record? (HINTS)
Perceived usefulness	In general, how useful is your online medical record for monitoring your health? (HINTS)
Intention to use	Was denken Sie heute, werden Sie sich Ihre medizinischen Daten und Unterlagen mit Hilfe der Karte zukünftig näher anschauen? (What do you think today, will you take a closer look at your medical data and documents with the help of the card in the future?) (Gesundheitsmonitor, Germany)
Preference to provide access to others	In order to get a quick and valid diagnosis, I am positive about giving internet access to my medical record to a doctor in another location or abroad (Poland)
Behavior	
Access frequency	Have you approached your family doctor, specialist, or other health professional(s) over the internet to read your health record? (Poland)
Type of information contents sought	귀하의 온라인 의료 기록에 다음과 같은 의료 정보가 포함되어 있습니까? (Do any of your online medical records include the following types of medical information?) (survey of CHISB[c])
Purpose of seeking a record	In the past 12 months, have you used your web-based medical record to…(look up test results, monitor your health, etc) (HINTS).
Channel	
Offline	
Attitude	
Perceived credibility	Do you believe health-related information from medical staff at medical centers or pharmacies? (Taiwan)
Perceived ease of use	How difficult is it to contact a doctor or other health care providers at this place after their regular hours in case of urgent medical needs-very difficult, somewhat difficult, not too difficult, or not at all difficult? (HTHS)
Satisfaction with service quality	How satisfied are you with the health care you received in the past 12 months? (NHIS[d])
Behavior	
Access frequency	How many times have you personally been to the doctor within the last 12 months (Europe)
Type of health service	What kind of place do you go to most often - a clinic, doctor's office, emergency room, or some other place? (NHIS)

XSL•FO
RenderX

Domain, subdomain, dimension, theme	Questionnaire items
Communication with health care provider	In the past 12 months, did health care provider talk with you about all of the different prescription medicines you are using, including medicines prescribed by other doctors? (HTHS[e])
Health related decision making	The following questions are about your communication with all doctors, nurses, or other health professionals you saw during the past 12 months: did they involve you in decisions about your health care as much as you wanted (HINTS)?
Online	
Attitude	
Perceived credibility	Selon vous, l'information de santé que vous avez obtenue la dernière fois est-elle crédible? (In your opinion, is the health information credible you obtained the last time (on the internet?) (France)
Perceived ease of use	In general, how comfortable do you feel. (using computers, internet, etc) (ANHCS[f])
Perceived usefulness	How useful was the health information you found online? (HTHS)
Perceived eHealth literacy	I know how to use the internet to answer my health questions (Israel)
Satisfaction with web-based information	Overall, how satisfied or not are you with the health-related information you found on the internet? (Europe)
Perceived confidentiality risks	There are different reasons for not approaching your family doctor, specialist, or other health professional(s) via the internet. Which reasons apply to you? (I worry about confidentiality) (Poland)
Intention to use	Next time you want to get information on health-related questions, how likely are you to use the internet? (Europe)
Behavior	
Access frequency	Within the last 12 months, have you used the internet to search for health-related information? (Europe)
Type of information technology device	Please indicate if you have each of the following: tablet computer like an iPad, smartphone, etc? (HINTS)
Health-related web and app (software use)	What kind of health apps do you currently have on your phone? (HTS)
Web-based resource (governmental website, Wikipedia, etc)	Have you used any of the following internet resources for health information? (government websites, news sites, etc) (ANHCS)
Communication with health care provider	Haben Sie diese Gesundheits-Apps auf Ihrem Tablet oder Smartphone schon einmal dazu genutzt, … um auf Gespräche mit Ihrem Arzt, Heilpraktiker, Physiotherapeuten usw. besser vorbereitet zu sein? (Have you ever used these health apps on your tablet or smartphone...to be better prepared for discussions with your doctor, alternative practitioner, physiotherapist, etc? (HINTS Germany)
Communication with friends and others (social media, forum, etc)	Still thinking just about the last 12 months, have you posted a health-related question online or shared your own personal health experience online in any way? (HTS)
Health-related decision making	Haben Sie diese Gesundheits-Apps auf Ihrem Tablet oder Smartphone schon einmal dazu genutzt, …. um zu entscheiden, wie mit einer Erkrankung umgegangen werden sollte? (Has your tablet or smartphone…helped you make a decision about how to treat an illness or condition? (HINTS Germany)
Tracking/managing health state	Has your tablet or smartphone helped you track progress on a health-related goal such as quitting smoking, losing weight, or increasing physical activity? (HINTS)
Improvement of health knowledge	Improved your understanding of the symptoms, conditions, or treatments in which you were interested (Israeli survey)
Health	
Overall health	
Attitude	
Perceived health efficacy	Overall, how confident are you about your ability to take good care of your health? (HINTS)
Concerns and belief about health	Agree that my good health is largely a matter of good fortune (ANHCS)
Behavior	
General health state	How would you rate your level of health in general? (Europe)

XSL•FO
RenderX

Domain, subdomain, dimension, theme	Questionnaire items
Diseases diagnosed	Are you now living with any of the following health problems or conditions (diabetes, high blood pressure, etc) (HTS)
Height	How tall are you without shoes? (NHIS)
Weight	About how much do you weigh, in pounds, without shoes? (HINTS)
Mental health	Have you been diagnosed with any of the following medical conditions? (mental health condition) (ANHCS)
Caregiving	Are you a caregiver for an adult family member with any of the following medical conditions? (Alzheimer disease, cancer, etc) (ANHCS)
Social support	Is there anyone you can count on to provide you with emotional support when you need it, such as talking over problems or helping you make difficult decisions? (HINTS)
Lifestyle	
Attitude	
Perception about nutrition	How likely is it that eating 5 or more servings of fruits and vegetables every day will (make you look better)? (ANHCS)
Perception about physical activity	How likely is it that doing at least moderate exercise 3 or more times a week will (reduce your feelings of stress)? (ANHCS)
Perception about alcohol	How much do you agree or disagree with each of the following statements? (alcohol increases your risk of cancer) (HINTS)
Perception about tobacco	In your opinion, do you think that some smokeless tobacco products such as chewing tobacco, snus, and snuff are less harmful to a person's health than cigarettes? (HINTS)
Behavior	
Nutrition	In the past week, on average, how many servings of fruit did you eat or drink per day? Please include 100% fruit juice, and fresh, frozen or canned fruits. (ANHCS)
Physical activity	In a typical week, how many days do you do any physical activity or exercise of at least moderate intensity, such as brisk walking, bicycling at a regular pace, and swimming at a regular pace? (HINTS)
Alcohol	In your entire life, have you had at least 12 drinks of any type of alcoholic beverage? (NHIS)
Tobacco	Have you smoked at least 100 cigarettes in your entire life? (ANHCS)
Cancer	
Attitude	
Perception about cancer	귀하께서는 다음 문항에 얼마나 동의하십니까? … 일상에서 접하는 모든 것이 암을 유발하는 원인임 (How much do you agree or disagree with each of the following statements? … It seems like everything causes cancer, There's not much you can do to lower your chances of getting cancer, etc) (survey of CHISB)
Behavior	
Cancer check-up	When did you have your most recent prostate-specific antigen test to check for prostate cancer? (ANHCS)
Cancer diagnosed	Have you ever been told by a doctor or other health professional that you had cancer or a malignancy of any kind? (NHIS)

[a]HINTS: Health Information National Trends Survey.

[b]HTS: Health Tracking Survey.

[c]CHISB: cancer and health-related information–seeking behavior.

[d]NHIS: National Health Interview Survey.

[e]HTHS: Health Tracking Household Survey.

[f]ANHCS: Annenberg National Health Communication Survey.

Channel

The channel can be defined as the means-focused domain that enables seekers to acquire and transmit health information [50]. The contents of the questionnaires pointed out that there were 2 channels for HISB: offline and online. The offline channel includes any method that collects or transmits health information through non–web-based sources such as health care providers, books, magazines, friends, seminars, or other means, and the offline subdomain consists of 7 themes (Figure 3). The online channel refers to seeking health information via the internet

XSL•FO
RenderX

with any information technology device; the online subdomain showed the largest number of themes, that is, 7 attitude and 9 behavior themes. The subdomains offline and online revealed a similar occurrence, with average percentages at 50.5% (SD 15.9) and 50.5% (SD 19.7), respectively. In particular, the average percentage of a behavioral dimension of the online channel, namely, access frequency, was counted as 92.3% in the selected HISB tools as well as 84.6% of the access frequency theme in the offline subdomain. The occurrence of behavior dimensions was 1.4-1.5 times that of the attitude dimensions.

Health

The health domain refers to the seeker's physical status and perceptions about health: overall health, lifestyle, and presence of cancer. Overall health refers to general health status, including physical, mental, and social health and concerns or beliefs about them. Lifestyle consists of 4 parts of a person's behavior and attitude: nutrition, physical activity, alcohol consumption, and tobacco consumption. Cancer themes focused on check-up and diagnosis with cancer perceptions. Overall health was the most frequently found subdomain out of the 7 subdomains (average percentage 53.8%, SD 18.0). The other subdomains, namely, lifestyle and cancer, revealed an average percentage of 32.7% and 46.2% with SD 21.3 and SD 13.3, respectively (Figure 3). In particular, the average percentage of behavioral themes in overall health and general health state accounted for 84.6% in the selected HISB tools, while perceptions about nutrition and physical activity accounted for the smallest percentage at 7.7%. Similar to other domains, the occurrence of behavior dimensions on the domain was 1.3-3.3 times higher than those of attitude.

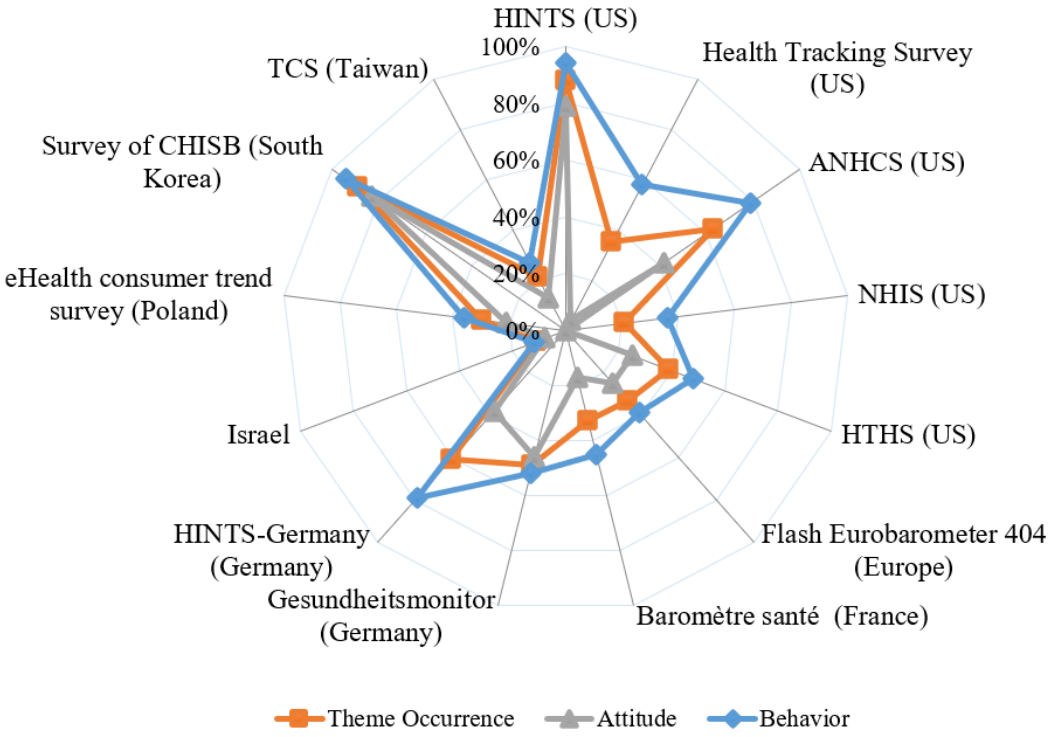
Person Characteristics

A person is the subject of HISB who seeks and utilizes information. A person's characteristics may affect HISB. The main considered characteristics throughout the instruments were age, sex or gender, nationality, race, language, education, income, occupational status, marital status, health literacy, health insurance, the number of household members, households with internet access, and preference for online or offline channels.

Themes Addressed by the Tools

The spider web diagram shows the average percentage of the themes in the selected HISB tools. Survey of CHISB (South Korea) and HINTS (United States) accounted for 89% (51/57) and 88% (50/57), respectively, which were found to be high average percentages among the tools (Figure 4). ANHCS (United States), HINTS Germany, and Gesundheitsmonitor (Germany) also contain 63% (36/57), 61% (35/57), and 49% (28/57) of the contents of HISB, respectively. Other tools including the HTHS (United States), NHIS (United States), the Health Tracking Survey (United States), the Flash Eurobarometer (Europe), Baromètre santé (France), the eHealth Consumer Trend Survey (Poland), and Taiwan Communication Survey (Taiwan) showed similar percentages of 21%-39% (12-22 out of 57 themes). The other HISB measurement from Israel showed only 11% (6/57) of the contents. All the tools focused more, by far, on the behavioral dimension than on attitude, showing a total average percentage of 53.8% and 30.4%, respectively; moreover, each average percentage of the behavior dimension accounted for 1.2-14.5 times more than the attitude throughout the instruments.

Figure 4. Average percentage of theme occurrence in health information–seeking behavior instruments. ANHCS: Annenberg National Health Communication Survey; HINTS: Health Information National Trends Survey; HISB: health information–seeking behavior; HTHS: Health Tracking Household Survey; NHIS: National Health Interview Survey; CHISB: cancer and health-related information–seeking behavior; TCS: Taiwan Communication Survey.



XSL•FO
RenderX

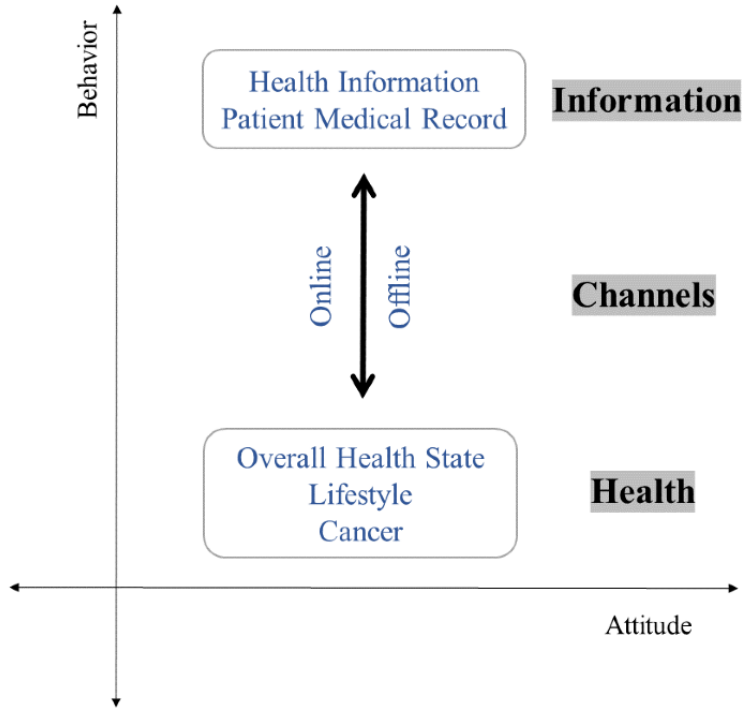
Sample Questionnaire Items

From the content analysis, representative sample questionnaire items from the 13 survey instruments were selected. Table 3 presents each questionnaire according to the domains, subdomains, and themes with attitude and behavior dimensions.

Proposed Theoretical Construct for Assessing HISB

Through the content analysis, a theoretical framework emerged. This study proposed the information-channel-health structure for assessing HISB (Figure 5). The theoretical structure shows reciprocal interaction between information and health through channels within the attitude and behavior dimensions. The information-channel-health concepts include the following: information, with information about health and patient medical records; channels, as online and offline; and health, with overall health state, lifestyle, and cancer. With the reciprocal structure of information-channel-health underlying 2 dimensions (attitude and behavior), the HISB phenomenon could be well illustrated with a comprehensive and holistic view.

Figure 5. A proposed theoretical construct for health information–seeking behavior.



Discussion

Principal Results

In this study, we investigated the main characteristics of the methodologies and the contents of the HISB survey tools used for over more than a decade (2008-2020) to answer the following research question: what are the characteristics of the measurement tools for assessing HISBs in nationally representative surveys around the world? The aim of this paper is to provide insights on the methodologies and the construct of content for HISB survey instruments from nationally representative studies. Through the systematic search, 13 survey tools were found in 2333 records related to HISB surveys. The features of this study's results are comprehensive and not limited to specific countries and specific topics or issue-based research. Other HISB-related review studies reported specific data such as age, college enrollment, adulthood, needs, and disease, including adolescent disease [82-84]. However, in this study, the results of the analysis were based on a tool for surveying healthy adults, who account for the highest proportion of the population density. Such a tool can lead to changes in the national policy.

The United States was found to have the most influential survey; 5 out of 13 tools developed in various countries were included in this study, a total of 188 research papers used data from HINTS, and HINTS identified 88% (50/57 themes) of the constructs, according to the findings. These strong features might be related to the purpose of HINTS to investigate respondents' access to and use of health information, including information technology to manage health and health information. The composition of most of the questionnaire tools was continuously updated according to the change of the cycle. However, in the current survey of HINTS 5 Cycle 4, researchers changed its scope to focus on cancer compared to prior HINTS surveys, which focused on health and medical topics. Therefore, HINTS 5 Cycle 3 was included for the contents analysis part of this study. In particular, owing to the influence of COVID-19, the questionnaire in France was changed twice in 2020 only to reduce the time of survey completion.

The contents of each country's survey tools contain the construct of HISBs. They can be categorized as information (information about health and patient medical records), channel (online and offline), and health (overall health state, lifestyle, and cancer), with dimensions of behavior and attitude. The questions are organized with more of the behavior dimension (average percentage of 53.8%) than attitude (average percentage of 30.4%) (Table 2). The analysis of the survey questionnaire contents conceptualized the HISB phenomenon, showing 3

XSL•FO
RenderX

domains, namely, information, channel, and health, with 2 dimensions, namely, behavior (objective outcome) and attitude (subjective tendency), emerging from the information-channel-health structure (Figure 5). In recent years, research has been conducted in parallel with existing reviews and meta-analysis to bring a theoretical framework to make some corrections [20] or to compare only specific variables to analyze the relationships [85] deductively. This study is meaningful as it derives a theoretical framework inductively after analyzing the contents by reviewing all the items of survey questionnaires. The findings of this study revealed that nationally representative surveys of HISB did not report theoretical frameworks when constructing the questionnaires. Therefore, it is believed that the outcomes of this study can be helpful in developing HISB-related tools or in establishing a theoretical framework prior to a large-scale investigation. This study included comprehensive (online and offline) HISBs. Recently, the terms eHealth and mobile health have become popular as many people use the internet and mobile access to manage their health. Therefore, preliminary review studies have focused on web-based HISBs or eHealth [17]. This research trend has a limitation in that it fails to address offline sources or face-to-face HISBs that still account for a large portion of HISBs.

This study found that all the survey instruments were from high-income countries, that is, United States, European Union, France, Germany, Israel, Poland, South Korea, and Taiwan, of the Organization for Economic Co-operation and Development [86]. The results can be interpreted as showing that there is information inequality, which may lead to a worsening of health inequality between high-income and low- and-middle-income countries. While low- and-middle-income countries still prioritize the establishment of universal health coverage focusing on the provider, high-income countries acknowledge the health information for individuals, empowering the health care consumer. The gap might be overcome through assessment of the trend of HISB in low- and-middle-income countries to contribute to the effective and efficient health care service to be provided. The details were analyzed by reviewing individual questions for the 13 survey tools, which were deeply rooted in the countries' differences. There are deviations in the questions according to the culture or medical system. For example, the question options vary depending on whether the countries are exposed to terrorism or have specific diseases or causes of cancer such as ultraviolet radiation exposure followed by a high incidence of skin cancer. In addition, questions about the type of health insurance and Medicare system also varied—for example, whether to visit in-store retail clinics, where to receive prescriptions, differences in the quality of and satisfaction with medical services, and accessibility to medical services.

The degree of information technology development in the country also has a great influence on the questions. The question asking whether the respondent has computers or mobile/smart devices depends on the development of information technology and the retention rate of mobile phones in each country. As an extension of this question, questions were subdivided into digital literacy, the type of fitness app, and whether web-based chat groups were used for health-related topics. With HINTS as a standard, related studies from Germany, South Korea, and China were also developed. HINTS Germany was established by HINTS (United States) and supported by the National Cancer Institute. In the case of South Korea, an individual researcher developed the survey questionnaires based on the content of HINTS and was funded by a national institute. HINTS China was excluded in this study because researchers did not conduct a nationally representative sample survey of the country. These studies would enable cross-national trend analysis and agenda for HISB.

Limitations and Recommendations for Future Research

For this study, we used databases in English and Korean, but there are some survey instruments that are neither English nor Korean. To overcome this limitation, we did not limit the languages in the search process. Moreover, it is obvious that English is the universal language of publication in the research field in the era of globalization. Therefore, we also used surveys in other languages, including 1 from France (French), Germany (German), and South Korea (Korean) in this paper. Some full versions of HISB survey instruments were not available for the review process. To attain the instrument, the researchers emailed corresponding authors for the HISB survey tools; however, these were found to be not related to HISB, or the author refused to provide a full version, or we received no response. In addition, the duration of the literature search was restricted to the period between 2008 and 2020. However, we mitigated this limitation because this study's findings cover the fundamental essence of HISB phenomena by analyzing existing tools over a more extensive period. The theoretical framework derived from this study could be used as a guide for nationally representative HISB surveys. From the findings of this study, we see that there was a lack of theoretical basis for the survey instrument. The framework including both the behavior/attitude and online/offline dimensions would provide integrative scope for national HISB phenomena. Moreover, this framework could be compared to other HISB-related theories, thereby enabling more comprehensive insight into the HISB phenomenon. As the study scope focused on HISB instruments that seek nationally representative samples, future studies could also analyze different populations, including certain regions, ages, genders, and occupations with HISB instruments. It would be worthy to compare the differences among the populations.

Conclusion

This study analyzed and synthesized current HISB survey questionnaires for nationally representative surveys. The findings of the methodology and content analysis provide a map and prototype for developing HISB-related instruments. A theoretical framework including both behavior/attitude and online/offline dimensions may provide integrative insight into real-world HISB phenomena. In sum, the findings of this study may contribute to better understanding of comprehensive HISB trends in nationally representative surveys.

XSL•FO
RenderX

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF-2019R1F1A1058969, 2017R1D1A1B03035510) and Nambu University, 2020.We thank Jamie Conklin (Health Sciences Librarian at University of North Carolina at Chapel Hill) who helped in the search process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies used for the study.
[PDF File (Adobe PDF File), 208 KB - jmir_v23i7e27539_app1.pdf]

Multimedia Appendix 2
List of studies related to each health information–seeking behavior survey instrument tool.
[PDF File (Adobe PDF File), 172 KB - jmir_v23i7e27539_app2.pdf]

Multimedia Appendix 3
Characteristics of the health information–seeking behavior survey instruments.
[PDF File (Adobe PDF File), 191 KB - jmir_v23i7e27539_app3.pdf]

Multimedia Appendix 4
Theme occurrence table.
[PDF File (Adobe PDF File), 241 KB - jmir_v23i7e27539_app4.pdf]

References

1. COVID-19 as a public health emergency of international concern (PHEIC) under the IHR. World Health Organization. URL: https://extranet.who.int/sph/covid-19-public-health-emergency-international-concern-pheic-under-ihr [accessed 2021-01-04]
2. Hernández-García I, Giménez-Júlvez T. Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717. [doi: 10.2196/18717] [Medline: 32217507]
3. Moreno A, Fuentes-Lara C, Navarro C. Covid-19 communication management in Spain: Exploring the effect of information-seeking behavior and message reception in public's evaluation. EPI 2020 May 29;29(4):e290402. [doi: 10.3145/epi.2020.jul.02]
4. Call for action: Managing the infodemic. World Health Organization. URL: https://www.who.int/news/item/11-12-2020-call-for-action-managing-the-infodemic [accessed 2021-01-04]
5. Associated Press. Coronavirus: In Iran, the false belief that toxic methanol fights Covid-19 kills hundreds. South China Morning Post. 2020 Mar 27. URL: https://www.scmp.com/news/world/middle-east/article/3077284/coronavirus-iran-false-belief-toxic-methanol-fights-covid-19 [accessed 2021-01-01]
6. Kim HK, Ahn J, Atkinson L, Kahlor LA. Effects of COVID-19 Misinformation on Information Seeking, Avoidance, and Processing: A Multicountry Comparative Study. Science Communication 2020 Sep 13;42(5):586-615. [doi: 10.1177/1075547020959670]
7. Park C. Coronavirus: saltwater spray infects 46 church-goers in South Korea. South China Morning Post. URL: https://www.scmp.com/week-asia/health-environment/article/3075421/coronavirus-salt-water-spray-infects-46-church-goers [accessed 2021-01-04]
8. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
9. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
10. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
11. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
12. Shaffer KM, Nightingale CL. Comparison of Healthcare Utilization Between Informal Caregivers and Non-Caregivers: An Analysis of the Health Information National Trends Survey. J Aging Health 2020;32(5-6):453-461 [FREE Full text] [doi: 10.1177/0898264319830262] [Medline: 30793639]

XSL•FO
RenderX

13. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US Population by Patterns of Mobile Health Use for Health and Behavioral Tracking: Analysis of the National Cancer Institute's Health Information National Trends Survey Data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]
14. O'Donnell JM, Jelinek GA, Gray KM, De Livera A, Brown CR, Neate SL, et al. Therapeutic utilization of meditation resources by people with multiple sclerosis: insights from an online patient discussion forum. Inform Health Soc Care 2020 Oct 01;45(4):374-384. [doi: 10.1080/17538157.2020.1755975] [Medline: 32508186]
15. Calixte R, Rivera A, Oridota O, Beauchamp W, Camacho-Rivera M. Social and Demographic Patterns of Health-Related Internet Use Among Adults in the United States: A Secondary Data Analysis of the Health Information National Trends Survey. Int J Environ Res Public Health 2020 Sep 19;17(18):6856 [FREE Full text] [doi: 10.3390/ijerph17186856] [Medline: 32961766]
16. Mano R. Use of Mobile Health Applications and the Self-Management of Cancer: A Gendered Approach. OJPM 2019;09(10):126-139. [doi: 10.4236/ojpm.2019.910012]
17. Hong Y, Cho J. Assessment of eHealth behaviors in national surveys: a systematic review of instruments. J Am Med Inform Assoc 2018 Dec 01;25(12):1675-1684 [FREE Full text] [doi: 10.1093/jamia/ocy128] [Medline: 30357344]
18. Bigsby E, Hovick SR. Understanding Associations between Information Seeking and Scanning and Health Risk Behaviors: An Early Test of the Structural Influence Model. Health Commun 2018 Mar;33(3):315-325. [doi: 10.1080/10410236.2016.1266575] [Medline: 28059570]
19. Phillips A. Proper Applications for Surveys as a Study Methodology. West J Emerg Med 2017 Jan;18(1):8-11 [FREE Full text] [doi: 10.5811/westjem.2016.11.32000] [Medline: 28116000]
20. Wang X, Shi J, Kong H. Online Health Information Seeking: A Review and Meta-Analysis. Health Commun 2020 Apr 14:1-13. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
21. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
22. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Global Health Data Exchange. URL: https://ghdx.healthdata.org/ [accessed 2021-02-17]
25. Health information national trends survey (HINTS) 5, cycle 3. National Cancer Institute. URL: https://hints.cancer.gov/data/survey-instruments.aspx [accessed 2021-02-17]
26. Health tracking survey. Pew Research Center. 2012. URL: https://www.pewresearch.org/internet/dataset/september-2012-health-tracking-prelim/ [accessed 2021-02-17]
27. Annenberg National Health Communication Survey (ANHCS). URL: https://anhcs.asc.upenn.edu/ [accessed 2021-02-17]
28. National health interview survey (NHIS). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm [accessed 2021-02-17]
29. Health tracking household survey (HTHS). Health and Medical Care Archive. 2010. URL: https://www.icpsr.umich.edu/web/HMCA/studies/34141 [accessed 2021-02-17]
30. Flash eurobarometer 404 (European citizen's digital health literacy). EU Open Data Portal. 2014. URL: https://data.europa.eu/data/datasets/s2020_404?locale=en [accessed 2021-02-17]
31. French health barometer (Baromètre santé). Sante Publique France. 2017. URL: https://www.santepubliquefrance.fr/etudes-et-enquetes/barometres-de-sante-publique-france/barometre-sante-2017 [accessed 2021-02-17]
32. Gesundheitsmonitor (Bertelsmann Healthcare Monitor). 2015. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/imported/abstract/1_680_Abstract.pdf [accessed 2021-02-17]
33. Hayat TZ, Brainin E, Neter E. With Some Help From My Network: Supplementing eHealth Literacy With Social Ties. J Med Internet Res 2017 Mar 30;19(3):e98 [FREE Full text] [doi: 10.2196/jmir.6472] [Medline: 28360024]
34. Bujnowska-Fedak MM. Trends in the use of the Internet for health purposes in Poland. BMC Public Health 2015 Feb 27;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
35. Stiftung Gesundheitswissen. HINTS Germany. 2019. URL: https://www.stiftung-gesundheitswissen.de/projekt-hints-germany [accessed 2021-02-17]
36. Hanna C. Classification of health information-seeking behavior among Korean adults. 2020 Presented at: The 23rd East Asian Forum of Nursing Scholars (EAFONS); Jan 10-11, 2020; Chiang Mai, Thailand.
37. Taiwan communication survey (TCS). Ministry of Science and Technology. 2016. URL: https://www.crctaiwan.nctu.edu.tw/annualsurvey_e.asp [accessed 2021-02-17]
38. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]

XSL•FO
RenderX

39. Beckjord EB, Finney Rutten LJ, Arora NK, Moser RP, Hesse BW. Information processing and negative affect: evidence from the 2003 Health Information National Trends Survey. Health Psychol 2008 Mar;27(2):249-257. [doi: 10.1037/0278-6133.27.2.249] [Medline: 18377144]
40. Shim M. Connecting internet use with gaps in cancer knowledge. Health Commun 2008 Sep;23(5):448-461. [doi: 10.1080/10410230802342143] [Medline: 18850392]
41. Van Stee SK, Yang Q. Online Cancer Information Seeking: Applying and Extending the Comprehensive Model of Information Seeking. Health Commun 2018 Dec;33(12):1583-1592. [doi: 10.1080/10410236.2017.1384350] [Medline: 29083231]
42. Chisolm DJ. Does online health information seeking act like a health behavior?: a test of the behavioral model. Telemed J E Health 2010 Mar;16(2):154-160. [doi: 10.1089/tmj.2009.0102] [Medline: 20156127]
43. Chisolm DJ, Sarkar M. E-health use in african american internet users: can new tools address old disparities? Telemed J E Health 2015 Mar;21(3):163-169. [doi: 10.1089/tmj.2014.0107] [Medline: 25536065]
44. Ardito S. Seeking consumer health information on the internet. Information Today Inc. 2013. URL: https://www.infotoday.com/OnlineSearcher/Articles/Medical-Digital/Seeking-Consumer-Health-Information-on-the-Internet-90558.shtml [accessed 2021-06-23]
45. Kuehn BM. More than one-third of US individuals use the Internet to self-diagnose. JAMA 2013 Feb 27;309(8):756-757. [doi: 10.1001/jama.2013.629] [Medline: 23443421]
46. Oh YS, Cho Y. Examining the relationships between resources and online health information seeking among patients with chronic diseases and healthy people. Soc Work Health Care 2015;54(2):83-100. [doi: 10.1080/00981389.2014.987940] [Medline: 25674723]
47. Ruggiero KJ, Gros DF, McCauley J, de Arellano MA, Danielson CK. Rural adults' use of health-related information online: data from a 2006 National Online Health Survey. Telemed J E Health 2011 Jun;17(5):329-334 [FREE Full text] [doi: 10.1089/tmj.2010.0195] [Medline: 21524201]
48. Saulsberry L, Price M, Hsu J. Medicare Medicaid Res Rev 2014;4(4):mmrr2014-004-04-b01 [FREE Full text] [doi: 10.5600/mmrr.004.04.b01] [Medline: 25383242]
49. Stern MJ, Cotten SR, Drentea P. The Separate Spheres of Online Health. Journal of Family Issues 2011 Oct 21;33(10):1324-1350. [doi: 10.1177/0192513X11425459]
50. Sun Y, Liu M, Krakow M. Health e-mavens: identifying active online health information users. Health Expect 2016 Oct;19(5):1071-1083 [FREE Full text] [doi: 10.1111/hex.12398] [Medline: 26296041]
51. Hovick SR, Bigsby E. Heart Disease and Colon Cancer Prevention Beliefs and Their Association With Information Seeking and Scanning. J Health Commun 2016;21(1):76-84. [doi: 10.1080/10810730.2015.1049307] [Medline: 26444664]
52. Ishikawa Y, Kondo N, Kawachi I, Viswanath K. Are socioeconomic disparities in health behavior mediated by differential media use? Test of the communication inequality theory. Patient Educ Couns 2016 Nov;99(11):1803-1807. [doi: 10.1016/j.pec.2016.05.018] [Medline: 27349600]
53. Kim S. An exploratory study of inactive health information seekers. Int J Med Inform 2015 Feb;84(2):119-133 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.003] [Medline: 25453277]
54. Lee CJ. Does the internet displace health professionals? J Health Commun 2008;13(5):450-464. [doi: 10.1080/10810730802198839] [Medline: 18661387]
55. Yang Q, Chen Y, Wendorf Muhamad J. Social Support, Trust in Health Information, and Health Information-Seeking Behaviors (HISBs): A Study Using the 2012 Annenberg National Health Communication Survey (ANHCS). Health Commun 2017 Sep;32(9):1142-1150. [doi: 10.1080/10410236.2016.1214220] [Medline: 27589249]
56. Li W, Watts J, Tan N. From Screen to Screening: Entertainment and News Television Media Effects on Cancer Screening Behaviors. J Health Commun 2019;24(4):385-394. [doi: 10.1080/10810730.2019.1607954] [Medline: 31033423]
57. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015 Apr 29;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
58. Dahlhamer JM, Galinsky AM, Joestl SS, Ward BW. Sexual Orientation and Health Information Technology Use: A Nationally Representative Study of U.S. Adults. LGBT Health 2017 Apr;4(2):121-129 [FREE Full text] [doi: 10.1089/lgbt.2016.0199] [Medline: 28287875]
59. US Centers For Disease Control And Prevention Epilepsy Program. Internet use and looking up information online in adults with epilepsy varies by epilepsy status--2013 National Health Interview Survey. Epilepsy Behav 2016 Jan;54:47-49. [doi: 10.1016/j.yebeh.2015.10.019] [Medline: 26655448]
60. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of Health Information Technology Use Between American Adults With and Without Chronic Health Conditions: Findings From The National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
61. Gonzalez M, Sanders-Jackson A, Wright T. Web-Based Health Information Technology: Access Among Latinos Varies by Subgroup Affiliation. J Med Internet Res 2019 Apr 16;21(4):e10389 [FREE Full text] [doi: 10.2196/10389] [Medline: 30990462]

XSL•FO
RenderX

62. Kindratt T, Callender L, Cobbaert M, Wondrack J, Bandiera F, Salvo D. Health information technology use and influenza vaccine uptake among US adults. Int J Med Inform 2019 Sep;129:37-42. [doi: 10.1016/j.ijmedinf.2019.05.025] [Medline: 31445279]
63. Heckman CJ, Handorf E, Auerbach MV. Prevalence and Correlates of Skin Cancer Screening Among Indoor Tanners and Nontanners. JAMA Dermatol 2018 May 01;154(5):554-560 [FREE Full text] [doi: 10.1001/jamadermatol.2018.0163] [Medline: 29617518]
64. Hong Y, Hincapie-Castillo JM, Xie Z, Segal R, Mainous AG. Socioeconomic and Demographic Characteristics of US Adults Who Purchase Prescription Drugs From Other Countries. JAMA Netw Open 2020 Jun 01;3(6):e208968 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8968] [Medline: 32579194]
65. Lee JH, Giovenco D, Operario D. Patterns of Health Information Technology Use according to Sexual Orientation among US Adults Aged 50 and Older: Findings from a National Representative Sample-National Health Interview Survey 2013-2014. J Health Commun 2017 Aug;22(8):666-671 [FREE Full text] [doi: 10.1080/10810730.2017.1341566] [Medline: 28749748]
66. Mahajan S, Lu Y, Spatz ES, Nasir K, Krumholz HM. Trends and Predictors of Use of Digital Health Technology in the United States. Am J Med 2021 Jan;134(1):129-134. [doi: 10.1016/j.amjmed.2020.06.033] [Medline: 32717188]
67. Yin R, Neyens DM. Online Health Resource Use by Individuals With Inflammatory Bowel Disease: Analysis Using the National Health Interview Survey. J Med Internet Res 2020 Sep 24;22(9):e15352 [FREE Full text] [doi: 10.2196/15352] [Medline: 32969831]
68. Rooks RN, Wiltshire JC, Elder K, BeLue R, Gary LC. Health information seeking and use outside of the medical encounter: is it associated with race and ethnicity? Soc Sci Med 2012 Jan;74(2):176-184. [doi: 10.1016/j.socscimed.2011.09.040] [Medline: 22154611]
69. Wiltshire JC, Roberts V, Brown R, Sarto GE. The effects of socioeconomic status on participation in care among middle-aged and older adults. J Aging Health 2009 Mar;21(2):314-335. [doi: 10.1177/0898264308329000] [Medline: 19091692]
70. Rooks RN, Kapral CG, Mathis AL. Chronic Conditions May Be More Important Than Race or Ethnicity in Relation to Health Information Seeking and Use. J Aging Health 2019 Apr;31(4):611-630. [doi: 10.1177/0898264317744643] [Medline: 29254425]
71. Dean CA, Geneus CJ, Rice S, Johns M, Quasie-Woode D, Broom K, et al. Assessing the significance of health information seeking in chronic condition management. Patient Educ Couns 2017 Aug;100(8):1519-1526. [doi: 10.1016/j.pec.2017.03.012] [Medline: 28320559]
72. Reifegerste D, Bachl M, Baumann E. Surrogate health information seeking in Europe: Influence of source type and social network variables. Int J Med Inform 2017 Jul;103:7-14. [doi: 10.1016/j.ijmedinf.2017.04.006] [Medline: 28551004]
73. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]
74. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
75. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. [How do people in Germany seek health information? Insights from the first wave of HINTS Germany]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Sep;63(9):1151-1160. [doi: 10.1007/s00103-020-03192-x] [Medline: 32666180]
76. Bujnowska-Fedak MM. Trends in the use of the Internet for health purposes in Poland. BMC Public Health 2015 Feb 27;15:194 [FREE Full text] [doi: 10.1186/s12889-015-1473-3] [Medline: 25886280]
77. Chang C. Self-Control-Centered Empowerment Model: Health Consciousness and Health Knowledge as Drivers of Empowerment-Seeking through Health Communication. Health Commun 2020 Nov;35(12):1497-1508. [doi: 10.1080/10410236.2019.1652385] [Medline: 31480856]
78. eHealth Consumer Trends Survey. 2007. URL: https://www.jmir.org/api/download?filename=a211f33fdca62ea09403094624688af0.pdf&alt_name=1023-5010-1-SP.pdf [accessed 2021-02-17]
79. Tove S. WHO/European eHealth consumer trends survey (final project report). Norwegian Center for eHealth Research. URL: https://ehealthresearch.no/en/reports/archive/who-european-ehealth-consumer-trends-survey [accessed 2021-02-17]
80. Rosenberg M, Hovland C. Cognitive, affective, and behavioral components of attitudes. In: McGuire W, Abelson R, Brehm J, editors. Attitude organization and change: An Analysis of Consistency Among Attitude Components. New Haven, CT: Yale University Press; 1960:1-14.
81. American Psychological Association (APA) Dictionary of Psychology. URL: https://dictionary.apa.org/ [accessed 2021-01-04]
82. Branscum P, Hayes L, Wallace L. Direct observation of searching for online health information: A systematic review of current evidence. American Journal of Health Studies. 2016. URL: https://www.amjhealthstudies.com/index.php/ajhs/article/view/154 [accessed 2021-06-22]
83. Pian W, Song S, Zhang Y. Consumer health information needs: A systematic review of measures. Information Processing & Management 2020 Mar;57(2):102077. [doi: 10.1016/j.ipm.2019.102077]
84. Kubb C, Foran HM. Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]

XSL•FO
RenderX

85. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
86. Doing business 2020: OECD high-income economies remain global benchmarks on most doing business indicators. The World Bank. URL: https://www.worldbank.org/en/news/press-release/2019/10/24/doing-business-2020-oecd-high-income-economies-remain-global-benchmarks-on-most-doing-business-indicators [accessed 2021-06-14]

Abbreviations

ANHCS: Annenberg National Health Communication Survey
CHISB: cancer and health-related information–seeking behavior
HINTS: Health Information National Trends Survey
HISB: health information–seeking behavior
HTHS: Health Tracking Household Survey
NHIS: National Health Interview Survey
RISS: Research Information Sharing Service

Edited by R Kukafka; submitted 28.01.21; peer-reviewed by Y Zhang, C Smith; comments to author 09.02.21; revised version received 20.04.21; accepted 06.05.21; published 26.07.21.

Please cite as:
Choi H, Jeong G
Characteristics of the Measurement Tools for Assessing Health Information–Seeking Behaviors in Nationally Representative Surveys: Systematic Review
J Med Internet Res 2021;23(7):e27539
URL: https://www.jmir.org/2021/7/e27539
doi:10.2196/27539
PMID:34309573

©Hanna Choi, Gyeonghui Jeong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Barriers to Dissemination of Local Health Data Faced by US State Agencies: Survey Study of Behavioral Risk Factor Surveillance System Coordinators

Manik Ahuja[1], PhD; Robert Aseltine Jr[2], PhD

[1]Department of Health Services Management and Policy, College of Public Health, East Tennessee State University, Johnson City, TN, United States
[2]Center for Population Health, UConn Health, Farmington, CT, United States

Corresponding Author:
Manik Ahuja, PhD
Department of Health Services Management and Policy
College of Public Health
East Tennessee State University
41B Lamb Hall
Johnson City, TN, 37604
United States
Phone: 1 4234396637
Email: ahujam@etsu.edu

Abstract

Background: Advances in information technology have paved the way to facilitate accessibility to population-level health data through web-based data query systems (WDQSs). Despite these advances in technology, US state agencies face many challenges related to the dissemination of their local health data. It is essential for the public to have access to high-quality data that are easy to interpret, reliable, and trusted. These challenges have been at the forefront throughout the COVID-19 pandemic.

Objective: The purpose of this study is to identify the most significant challenges faced by state agencies, from the perspective of the Behavioral Risk Factor Surveillance System (BRFSS) coordinator from each state, and to assess if the coordinators from states with a WDQS perceive these challenges differently.

Methods: We surveyed BRFSS coordinators (N=43) across all 50 US states and the District of Columbia. We surveyed the participants about contextual factors and asked them to rate system aspects and challenges they faced with their health data system on a Likert scale. We used two-sample *t* tests to compare the means of the ratings by participants from states with and without a WDQS.

Results: Overall, 41/43 states (95%) make health data available over the internet, while 65% (28/43) employ a WDQS. States with a WDQS reported greater challenges (*P*=.01) related to the cost of hardware and software (mean score 3.44/4, 95% CI 3.09-3.78) than states without a WDQS (mean score 2.63/4, 95% CI 2.25-3.00). The system aspect of standardization of vocabulary scored more favorably (*P*=.01) in states with a WDQS (mean score 3.32/5, 95% CI 2.94-3.69) than in states without a WDQS (mean score 2.85/5, 95% CI 2.47-3.22).

Conclusions: Securing of adequate resources and commitment to standardization are vital in the dissemination of local-level health data. Factors such as receiving data in a timely manner, privacy, and political opposition are less significant barriers than anticipated.

(J Med Internet Res 2021;23(7):e16750) doi:10.2196/16750

KEYWORDS

web-based data query systems, WDQS; health data; population health; dissemination of local health data

Introduction

It is widely acknowledged in the public health field that progress in information technology has paved the way for exciting opportunities to disseminate local level health data more efficiently [1,2]. The growth of the internet, mobile technologies, artificial intelligence, and other technological advances have enabled health information to become more easily accessible and widely available to a broader population [2-4]. There has

XSL•FO
RenderX

been growing enthusiasm for the application of big data and its utility in public health, particularly for population-level health data [5]. Access to high quality population-level health data is essential for public health, as it informs us of disease tracking, health problems, and health surveillance at the subpopulation level [6,7]. Health informatics has enabled public health practitioners to assess public and population health information by accurately combining data from a wide range of disparate sources [7]. Despite the vast advances in technology, there are many challenges associated with availability of high-quality population-level health data [7,8]. These problems have been even further exacerbated during the COVID-19 pandemic, as there is a lack of consistency in the data and their reporting [9,10].

As each US state is responsible for its own health surveillance, decisions regarding technology implementation have varied from state to state [11]. One popular option is for state agencies to design and develop dynamic web-based data query systems (WDQSs), which allow users to customize data queries by choosing data sets, variables, measures, and the format for presenting query results [12]. There are many benefits to WDQSs, as state agencies are able to respond to data requests in a timely manner, provide data to a broader population, and contribute to the development of community health assessments and policy decisions [13]. Another popular option is for states to make static reports available to the public on the World Wide Web. Static reports are generally manually compiled by staff and are available in formats such as PDF, Microsoft Excel, and HTML. Static reports limit user choice to precalculated statistics and do not allow users to choose parameters for a query [14]. Given the advances in information technology, deployment of these outdated technologies, which can be difficult to use and plagued by missing or incomplete data, by US states is unfortunate [15]. In the private sector, industries such as finance have been successful in maximizing the potential of the internet, as changes in stock prices are made available within seconds to end users.

WDQSs were first implemented in the late 1990s, as states developed systems in which queries could be specified and results returned on the World Wide Web without requiring any additional software [12]. At the time, strong efforts were being made to achieve data liberation and use of open-source information technology solutions and collaboration to promote public health [15]. Collaboration has contributed to states sharing developments, ideas, and knowledge to meet a variety of public health assessment needs [16-18]. To reduce the cost burden, there has been a push toward open-source software, which costs little or no money to procure [19]. Open-source code can be easily shared, and its key benefits include free redistribution, inclusion of source code, easy modification of the code, and lack of need for an additional license [20]. For example, in 2000, Utah's legacy Indicator Based Information System–Public Health (IBIS-PH) was developed using open-source code, and states such as New Mexico and Kentucky adopted the system within the first few years of its development. However, over the last decade, efforts toward WDQS implementation have stalled, and progress has slowed. Public health is at risk of falling behind from a technological standpoint [21]. Despite the significant advances in technology, including faster processors, improved bandwidth, and lower cost of storage, state agencies commonly face data sharing barriers between organizations within their state [21]. The purpose of this study was to investigate the barriers faced by state agencies to the implementation of WDQSs from the perspectives of Behavioral Risk Factor Surveillance System (BRFSS) coordinators from each state. We aimed to understand how challenges and perceptions of systems differ between states that have implemented a WDQS and states that have not. We hypothesized that the perceptions and challenges reported by the BRFSS coordinators would vary across states because of the coordinators' significance in releasing health data and their role in technology decisions in their states [14].

Methods

Study Design and Sampling

We designed and administered a web-based questionnaire to BRFSS coordinators from all 50 states and the District of Columbia (n=51). The BRFSS, established in 1984, is a health-related telephone survey that collects state data from US residents in all 50 states regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. The BRFSS coordinator in each state is responsible for gathering information about health behaviors and is responsible for the management and oversight of the BRFSS survey [22,23]. First, we queried the coordinators on how states disseminate their data. In January 2015, each prospective participant was sent an invitation letter by US Mail. We found the names and contact information of all the BRFSS coordinators on the US Centers for Disease Control and Prevention website [24]. The letter provided details of the study and indicated that participation was confidential, as the results would not be disclosed to anyone except the study staff. The letter specified that the coordinators' participation was voluntary and that they could withdraw at any point of the survey. Each participant was sent a follow-up email containing a link to the survey with a secure user name and password within 1-2 weeks upon receipt of the letter.

We queried the participants on the basic characteristics of their data dissemination, including how their health data are disseminated, which technology is used, and which types of data are available. We asked, "Does your state present health data over the internet?" and participants answered yes or no. We also queried if their state presents data using an interactive WBQS, which types of data are publicly available (eg, BRFSS, births, deaths), and finally, which path they took to develop their software (eg, in-house, adopted from an outside vendor). We also asked participants to rate the level of challenges they face regarding the dissemination of their data on a 4-point Likert scale (1, not at all challenging; 2, not very challenging; 3, somewhat challenging; 4, very challenging). A Likert scale is a set of statements (items) offered for a hypothetical situation under study, in which participants are asked to show their level of agreement (eg, strongly agree, somewhat agree). We presented 2 to 4 questions for three different categories: cost, staffing and support, and other challenges. We then asked the

XSL•FO
RenderX

coordinators how they would rate various system-related aspects of their health data systems. They were presented a list of items and rated each item on a 5-point Likert scale (1, poor; 2, fair; 3, good; 4, very good; 5, excellent). For the system aspects, 3 to 4 questions were presented for each category, including website performance, data quality, and accessibility and support.

Data Analysis

We obtained descriptive statistics and profile characteristics for the participants from each state. We also compared the means of the ratings of participants from states that have implemented a WDQS and from states that have not implemented a WDQS. The means were compared using two-sample *t* tests between states with and without a WDQS, with $P<.05$ used as the level of significance.

Ethics Approval and Consent to Participate

Ethics approval for the study protocol was received from the Human Subjects Protection Office at the University of Connecticut Health Center.

Results

Profile characteristics at the state level can be found in Table 1. A total of 43 of the 51 coordinators completed the survey, for an overall response rate of 84%. Of the 43 participants, 42 completed the web-based survey, while 1 participant completed the survey over the telephone. Of the 43 states, 41 (95%) had some form of health data available on the web, while 28 (65%) reported having implemented a WDQS. Data available included BRFSS (33/43, 77%), births (31/43, 72%), deaths (30/43, 70%), lead screening (5/43, 12%), and hospitalizations (4/43, 9%).

Table 1. Profile of health data characteristics at the state level (n=43).

Characteristic and responses	Value, n (%)
Health data are made available on the internet	
Yes	41 (95)
No	2 (5)
Health data are made available using an interactive web-based query system	
Yes	28 (65)
No	15 (35)
Types of data that are publicly available[a]	
BRFSS[b]	33 (73)
Births	31 (72)
Deaths	30 (70)
Lead screening	5 (12)
Hospitalization	4 (9)
Software development path	
In-house	13 (37)
Outside vendor	8 (19)
Adopted from another state	6 (14)
Off the shelf commercial software	4 (9)

[a]More than one response is acceptable for this question, as a state may have multiple data sources.

[b]BRFSS: Behavioral Risk Factor Surveillance System.

We report the mean ratings for the challenges faced and system aspects in Table 2. On a 4-point Likert scale, participants rated the cost of system development (mean score 3.33), consultants/vendors (mean score 3.05), and the cost of hardware/software (mean score 2.89) as the most challenging. The overall mean score for all challenges faced was 2.68 on the 4-point Likert scale. Participants reported the lack of political support (mean score 1.77) and issues with data privacy (mean 2.55) as less of a challenge. We analyzed the mean results of the reported challenges between states with and without a WDQS (Table 2). Participants from states without a WDQS reported the cost of hardware and software to be a greater challenge than those from states with a WDQS (mean score 3.55, 95% CI 3.09-3.78, vs mean score 2.63, 95% CI 2.25-3.00; $P=.01$). System aspects were rated higher ($P=.01$) for the standardization of vocabulary by participants from states with a WDQS (mean 3.32, 95% CI=2.94-3.69) versus those from states without a WDQS (mean=2.85, 95% CI 2.47-3.22).

XSL•FO
RenderX

Table 2. Participant ratings of challenges faced by state agencies and of system aspects (N=43).

Grouping and item	Overall mean score[a]	With WDQS[b] (n=28), mean (95% CI)	Without WDQS (n=15), mean (95% CI)	*P* value[c]
Challenges faced (scored on a 4-point Likert scale)				
Cost				
Cost of system development	3.33	3.05 (2.68-3.41)	3.6 (3.28-3.91)	.06
Cost of hardware/software	2.89	2.63 (2.25-3.00)	3.44 (3.09-3.78)	*.01* [d]
Cost of vendors/consultants	3.06	2.93 (2.44-3.41)	3.25 (2.63-3.86)	.42
Staffing and support				
Lack of internal information technology staff	2.97	2.90 (2.37-3.42)	3.10 (2.48-3.71)	.67
Help desk support	2.60	2.40 (1.80-3.02)	3.00 (2.25-3.74)	.27
Lack of trained staff who understand the data	2.62	2.42 (1.98-2.85)	3.00 (2.49-3.50)	.12
Receiving data in a timely manner	2.40	2.36 (2.07-2.76)	2.42 (1.98-2.85)	.85
Other challenges				
Privacy	2.55	2.37 (2.06-2.67)	2.9 (2.28-3.51)	.11
Political opposition	1.77	1.6 (1.10-2.11)	2.14 (1.47-2.80)	.23
System aspects (scored on a 5-point Likert scale)				
Website usability				
User-friendliness	3.15	3.15 (2.73-3.56)	3.00 (2.69-3.32)	.64
Website performance	3.34	3.54 (3.15-3.93)	2.92 (2.47-3.36)	.06
Standardization of vocabulary	3.14	3.32 (2.94-3.69)	2.85 (2.47-3.22)	*.01*
End user satisfaction	3.03	3.37 (2.89-3.84)	2.40 (2.08-2.71)	.11
Data quality				
Availability of race, gender, and other social determinants	3.91	4.05 (3.63-4.46)	3.64 (3.24-4.03)	.11
Quality of data	3.90	4.00 (3.62-4.37)	3.69 (3.28-4.09)	.33
Breadth of data	3.33	3.42 (2.96-3.87)	3.15 (2.71-3.58)	.46
Ability to link to multiple data sources	2.81	3.28 (2.57-3.98)	1.75 (1.26-2.24)	*.01*
Accessibility and support				
Accessibility to researchers	3.67	3.92 (3.56-4.27)	3.15 (2.66-3.63)	.06
Accessibility to nonresearchers	3.58	3.76 (3.39-4.12)	3.23 (2.69-3.76)	.11
Timeliness of support requests	3.53	3.68 (3.28-4.07)	3.27 (2.73-3.80)	.23

[a]The overall mean score represents the full sample.

[b]WDQS: web-based data query system.

[c]Two-sample *t* tests were used to compare the mean scores between states with and without a WDQS.

[d]Italic text indicates statistical significance at $P<.05$.

Discussion

To our knowledge, this is the first national study to investigate barriers faced by state agencies to the dissemination of their health data using informants in key roles. The findings revealed that BRFSS coordinators rated their systems more favorably in states where a WDQS was implemented. Interestingly, despite the high cost of technology, staffing, implementation, and maintenance of technology-based systems and other factors, BRFSS coordinators from states that implemented a WDQS perceived their systems more favorably. We hypothesize that these findings are indications of a favorable assessment of the cost-benefit ratio of implementation of technology-based systems relative to low-cost health data systems. Adequate staffing and funding for state health data systems is lacking, which has impeded or slowed progress or halted data dissemination efforts in these states [25,26]. Our findings are more important than ever, given the reliance of society on trusted, reliable, and accurate public health data [27,28].

Prior research has indicated that organizations are reluctant to share their data due to organizational, technical, and political barriers [15]. In the current study, respondents reported that

XSL•FO
RenderX

state agencies are generally willing to share their data and do not perceive political barriers as a significant challenge to data sharing. However, these findings should be taken with caution, as there may be bias because these perspectives were based exclusively on the experience of BRFSS coordinators, which may not be representative of that of other key stakeholders across states. Respondents also reported lack of interoperability between systems, as data may be transmitted in formats that are incompatible with the originating system. These findings are in line with prior work, in which it was reported that departments lack adequate staffing and resources to profile, "cleanse," and manipulate these data so they are usable [15,29]. If data are not usable, they have limited utility and do not create significant opportunities for public health research. According to the latest Public Health Workforce Interests and Needs Survey (PH WINS), a nationally representative survey of the public health workforce, state agencies lack adequate trained staff who are able to handle and interpret these data [30,31]. Public health agencies are lacking workers in areas such as data-informed decision-making, health informatics, and data quality, which are essential in the dissemination of public health data [30-32].

Our findings should be interpreted with certain limitations in mind. First, the results may not be generalizable beyond the perspective of the BRFSS coordinator in each state. As each state may have multiple stakeholders who have a vested interest in the WDQS, the opinion of the BRFSS coordinator may not be representative of the consensus from that state. Secondly, our study may reflect bias, as BRFSS coordinators in states with a WDQS may rate their systems higher due to the additional investments states have made in this technology. Third, because the study includes a small number of participants, there is insufficient statistical power to detect small differences in ratings among states with and without a WDQS. Fourth, as BRFSS coordinators from 7/51 states (14%) did not respond to the survey, there may be systematic bias related to the missing information from these states. The reasons that the BRFSS coordinators from those states refused to participate are also unknown. Fifth, questions may be interpreted differently from one state to the next. Web-based expertise and technical maturity may also vary from one state to another, depending on their experience. Finally, although measures such as quality, timeliness, satisfaction, and access were assessed for multiple constructs, their definitions were not presented in the survey. Respondents may have interpreted these measures differently, potentially resulting in bias. For example, the definition of "quality" may be perceived differently from one state to the next. Despite these limitations, the current study is, to our knowledge, the first to compare system ratings and assessments of challenges to presenting health data to the public among states with more primitive versus more advanced data systems. Directions for future research include more comprehensive efforts to evaluate the utility of WDQSs, as evidence of their usefulness and their potential impact on public health may help justify the additional expenditures required. Additionally, it is recommended that state agencies aim toward collaboration and investigate open-source software options. This model has been successful in the clinical setting. For example, open-source software has been adopted by several hospitals and clinics. A similar model can be applied for future WDQS development, as states should aim to collaborate and work toward building robust systems that are easy adoptable. In summary, it is important to design systems that facilitate access to local health data; these data provide information regarding health challenges at the subpopulation level, which will ultimately help guide future public health research. These problems have been at the forefront during the COVID-19 pandemic [33,34] and should be urgently addressed moving forward.

Data Availability

Data and materials are available upon request.

Authors' Contributions

MA is the lead author and contributed to the design, analysis, and writing. RA contributed to the design, analysis, and writing.

Conflicts of Interest

None declared.

References

1. Buehler JW, Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century. Introduction. MMWR Suppl 2012 Jul 27;61(3):1-2. [Medline: 22832989]
2. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]
3. Khan ZF, Alotaibi SR. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. J Healthc Eng 2020 Sep 01;2020:8894694-8894615 [FREE Full text] [doi: 10.1155/2020/8894694] [Medline: 32952992]
4. Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med 2018;1:53 [FREE Full text] [doi: 10.1038/s41746-018-0058-9] [Medline: 31304332]
5. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health 2019 Oct 01;29(Supplement_3):23-27 [FREE Full text] [doi: 10.1093/eurpub/ckz168] [Medline: 31738444]

XSL•FO
RenderX

6. Bakshi S, Schiavoni KH, Carlson LC, Chang TE, Flaster AO, Forester BP, et al. The essential role of population health during and beyond COVID-19. Am J Manag Care 2021 Mar 30;27(3):123-128 [FREE Full text] [doi: 10.37765/ajmc.2021.88511] [Medline: 33720669]
7. Massoudi BL, Chester KG. Public health, population health, and epidemiology informatics: recent research and trends in the United States. Yearb Med Inform 2017 Aug 19;26(01):241-247. [doi: 10.1055/s-0037-1606510]
8. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
9. Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. J Infect Public Health 2020 Nov;13(11):1630-1638 [FREE Full text] [doi: 10.1016/j.jiph.2020.07.019] [Medline: 32855090]
10. Sáez C, Romero N, Conejero J, García-Gómez JM. Potential limitations in COVID-19 machine learning due to data source variability: a case study in the nCov2019 dataset. J Am Med Inform Assoc 2021 Feb 15;28(2):360-364 [FREE Full text] [doi: 10.1093/jamia/ocaa258] [Medline: 33027509]
11. Gjelsvik A, Buechner J. States' approaches to implementing a web-based data query system. J Public Health Manag Pract 2006;12(2):170-175. [doi: 10.1097/00124784-200603000-00009] [Medline: 16479231]
12. Friedman DJ, Parrish RG. Characteristics, desired functionalities, and datasets of state web-based data query systems. J Public Health Manag Pract 2006;12(2):119-129. [doi: 10.1097/00124784-200603000-00003] [Medline: 16479225]
13. Shah G, Leep C, Alexander D. Profile-IQ: web-based data query system for local health department infrastructure and activities. J Public Heal Manag Pract 2014;20(2). [doi: 10.1097/phh.0b013e3182a1beb7]
14. Friedman DJ. Envisioning futures for state Web-based data query systems. J Public Health Manag Pract 2006;12(2):196-200. [doi: 10.1097/00124784-200603000-00013] [Medline: 16479235]
15. Love D, Shah GH. Reflections on organizational issues in developing, implementing, and maintaining state Web-based data query systems. J Public Health Manag Pract 2006;12(2):184-188. [doi: 10.1097/00124784-200603000-00011] [Medline: 16479233]
16. Kovach K, Welter C, Seweryn S, Torres G. Perceived benefits of collaboration between local health departments and schools and programs of public health: a mixed-methods study. J Public Heal Manag Pract 2019;25(2):A. [doi: 10.1097/phh.0000000000000823]
17. Valaitis RK, O'Mara L, Wong ST, MacDonald M, Murray N, Martin-Misener R, et al. Strengthening primary health care through primary care and public health collaboration: the influence of intrapersonal and interpersonal factors. Prim Health Care Res Dev 2018 Apr 12;19(04):378-391. [doi: 10.1017/s1463423617000895]
18. Shahzad M, Upshur R, Donnelly P, Bharmal A, Wei X, Feng P, et al. A population-based approach to integrated healthcare delivery: a scoping review of clinical care and public health collaboration. BMC Public Health 2019 Jun 07;19(1):708 [FREE Full text] [doi: 10.1186/s12889-019-7002-z] [Medline: 31174501]
19. Wheeler DA. Why open source software / free software (OSS/FS, FLOSS, or FOSS)? Look at the numbers!. Challenges 2014. URL: http://dwheeler.com/oss_fs_why.html [accessed 2021-07-05]
20. Aminpour F, Sadoughi F, Ahmadi M. Towards the application of open source software in developing national electronic health record-narrative review article. Iran J Public Health 2013 Dec;42(12):1333-1339 [FREE Full text] [Medline: 26060634]
21. Vest J, Issel LM, Lee S. Experience of using information systems in public health practice: findings from a qualitative study. Online J Public Health Inform 2014 Feb 05;5(3):227 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4847] [Medline: 24678380]
22. Anderson N, Marcum J. Using behavioral risk factor surveillance system data as an occupational health profile: Washington State janitors, 2011 to 2017. J Occup Environ Med 2019;61(9). [doi: 10.1097/jom.0000000000001652]
23. Iachan R, Pierannunzi C, Healey K, Greenlund KJ, Town M. National weighting of data from the Behavioral Risk Factor Surveillance System (BRFSS). BMC Med Res Methodol 2016 Nov 15;16(1):155 [FREE Full text] [doi: 10.1186/s12874-016-0255-7] [Medline: 27842500]
24. BRFSS State Coordinators. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/brfss/state_info/coordinators.htm [accessed 2021-07-06]
25. McFarlane T, Dixon B, Grannis S, Gibson P. Public health informatics in local and state health agencies: an update from the public health workforce interests and needs survey. J Public Health Manag Pract 2019:A. [doi: 10.1097/phh.0000000000000918]
26. Colquhoun A, Aplin L, Geary J, Goodman KJ, Hatcher J. Challenges created by data dissemination and access restrictions when attempting to address community concerns: individual privacy versus public wellbeing. Int J Circumpolar Health 2012 May 08;71:1-7 [FREE Full text] [doi: 10.3402/ijch.v71i0.18414] [Medline: 22584511]
27. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? Open data for healthcare. Front Public Health 2016;4:7 [FREE Full text] [doi: 10.3389/fpubh.2016.00007] [Medline: 26925395]
28. Edelstein M, Lee LM, Herten-Crabb A, Heymann DL, Harper DR. Strengthening global public health surveillance through data and benefit sharing. Emerg Infect Dis 2018 Jul;24(7):1324-1330. [doi: 10.3201/eid2407.151830]
29. Shah G, Madamala K. Knowing where public health is going: levels and determinants of workforce awareness of national public health trends. J Public Health Manag Pract 2015;21. [doi: 10.1097/phh.0000000000000312]

XSL•FO
RenderX

30. Bogaert K, Castrucci B, Gould E, Sellers K, Leider J, Whang C, et al. The public health workforce interests and needs survey (PH WINS 2017): an expanded perspective on the state health agency workforce. J Public Health Manag Pract 2019;25:A. [doi: 10.1097/phh.0000000000000932]
31. Yeager V, Balio C, Kronstadt J, Beitsch L. The relationship between health department accreditation and workforce satisfaction, retention, and training needs. J Public Health Manag Pract 2019;25. [doi: 10.1097/phh.0000000000000920]
32. Rodriguez H, Ramirez J. The work experiences of community health professionals: implications for the continued rollout of the affordable care act. J Public Health Manag Pract 2015;21. [doi: 10.1097/phh.0000000000000307]
33. Fegan G, Cheah PY. Solutions to COVID-19 data sharing. Lancet Digital Health 2021 Jan;3(1):e6. [doi: 10.1016/s2589-7500(20)30273-9]
34. Gao F, Tao L, Huang Y, Shu Z. Management and data sharing of COVID-19 pandemic information. Biopreserv Biobank 2020 Dec 01;18(6):570-580. [doi: 10.1089/bio.2020.0134] [Medline: 33320734]

Abbreviations

BRFSS: Behavioral Risk Factor Surveillance System
IBIS-PH: Indicator Based Information System–Public Health
PH WINS: Public Health Workforce Interests and Needs Survey
WDQS: web-based data query system

Edited by R Kukafka; submitted 21.10.19; peer-reviewed by A Jimenez, V Ramos, A Khurshid, E Hsu; comments to author 25.09.20; revised version received 19.02.21; accepted 24.05.21; published 13.07.21.

Please cite as:
Ahuja M, Aseltine Jr R
Barriers to Dissemination of Local Health Data Faced by US State Agencies: Survey Study of Behavioral Risk Factor Surveillance System Coordinators
J Med Internet Res 2021;23(7):e16750
URL: https://www.jmir.org/2021/7/e16750
doi:10.2196/16750
PMID:34255650

©Manik Ahuja, Robert Aseltine Jr. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

One Decade of Online Patient Feedback: Longitudinal Analysis of Data From a German Physician Rating Website

Martin Emmert[1*], MHMM, PhD; Stuart McLennan[2,3*], MBHL, PhD

[1]Institute for Healthcare Management & Health Sciences, University of Bayreuth, Bayreuth, Germany

[2]Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany

[3]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[*]all authors contributed equally

Corresponding Author:
Martin Emmert, MHMM, PhD
Institute for Healthcare Management & Health Sciences
University of Bayreuth
Prieserstraße 2
Bayreuth, 95444
Germany
Phone: 49 921 55 ext 4827
Email: martin.emmert@uni-bayreuth.de

Abstract

Background: Feedback from patients is an essential element of a patient-oriented health care system. Physician rating websites (PRWs) are a key way patients can provide feedback online. This study analyzes an entire decade of online ratings for all medical specialties on a German PRW.

Objective: The aim of this study was to examine how ratings posted on a German PRW have developed over the past decade. In particular, it aimed to explore (1) the distribution of ratings according to time-related aspects (year, month, day of the week, and hour of the day) between 2010 and 2019, (2) the number of physicians with ratings, (3) the average number of ratings per physician, (4) the average rating, (5) whether differences exist between medical specialties, and (6) the characteristics of the patients rating physicians.

Methods: All scaled-survey online ratings that were posted on the German PRW jameda between 2010 and 2019 were obtained.

Results: In total, 1,906,146 ratings were posted on jameda between 2010 and 2019 for 127,921 physicians. The number of rated physicians increased constantly from 19,305 in 2010 to 82,511 in 2018. The average number of ratings per rated physicians increased from 1.65 (SD 1.56) in 2010 to 3.19 (SD 4.69) in 2019. Overall, 75.2% (1,432,624/1,906,146) of all ratings were in the best rating category of "very good," and 5.7% (107,912/1,906,146) of the ratings were in the lowest category of "insufficient." However, the mean of all ratings was 1.76 (SD 1.53) on the German school grade 6-point rating scale (1 being the best) with a relatively constant distribution over time. General practitioners, internists, and gynecologists received the highest number of ratings (343,242, 266,899, and 232,914, respectively). Male patients, those of higher age, and those covered by private health insurance gave significantly (P<.001) more favorable evaluations compared to their counterparts. Physicians with a lower number of ratings tended to receive ratings across the rating scale, while physicians with a higher number of ratings tended to have better ratings. Physicians with between 21 and 50 online ratings received the lowest ratings (mean 1.95, SD 0.84), while physicians with >100 ratings received the best ratings (mean 1.34, SD 0.47).

Conclusions: This study is one of the most comprehensive analyses of PRW ratings to date. More than half of all German physicians have been rated on jameda each year since 2016, and the overall average number of ratings per rated physicians nearly doubled over the decade. Nevertheless, we could also observe a decline in the number of ratings over the last 2 years. Future studies should investigate the most recent development in the number of ratings on both other German and international PRWs as well as reasons for the heterogeneity in online ratings by medical specialty.

(J Med Internet Res 2021;23(7):e24229) doi:10.2196/24229

KEYWORDS

physician rating websites; patient satisfaction; patient feedback; online ratings

XSL•FO
RenderX

Introduction

Feedback from patients is an essential element of a patient-oriented health care system [1]. Patients' views and opinions on the care they have experienced can help health care organizations and professionals identify areas that need to be improved and can also help other patients with decision making when choosing where to receive health care [2]. Health care organizations and professionals can gather patient feedback in a variety of ways, including by conducting patient surveys, audits, interviews, focus groups, and deliberative events [3]. Patients have also always been able to actively share their views and opinions about the care they received with family and friends or with health care organizations and professionals via unsolicited comments or complaints. However, patients increasingly also have the ability to share their views and opinions on the internet and social media [4-7].

Physician rating websites (PRWs) are one of the key opportunities for patients to provide feedback online [4,7]. A systematic search of PRWs in 2018 identified 143 websites from 12 countries; however, the majority of websites were commercially operated in the United States and Germany [8]. Previous research involving PRW ratings in Germany and other countries has highlighted some common themes, including incomplete lists of physicians, a low number of physicians rated, a low number of ratings per physician that are overwhelmingly positive, and unstructured and different rating systems, which has raised concerns about the representativeness, validity, and usefulness of feedback on PRWs [7,9-30]. Medical associations have also often expressed strong opposition to PRWs, concerned that they will be used for doctorbashing or defamation [31,32]. Countries have different legal frameworks with regards to data protection, and previous research suggests that restrictive legal environments (eg, Switzerland) may be having an impact of the types of ratings on PRWs [28,29]. However, the legal basis for PRWs in Germany is reasonably liberal and well established. The Federal Court of Justice of Germany confirmed in 2014 the permissibility of ratings on the basis of the right to freedom of expression and that the anonymity of raters can only be lifted in exceptional cases [33,34]. Research also indicates that PRWs in Germany are having some success in influencing patient decision making and quality improvement [17,35].

However, most studies examining PRWs ratings have typically focused on a certain year (eg, [13,18,21]), a certain medical specialty (eg, [22,23,36-40]), certain cities or regions (eg, [14,26,41]), or with a (more or less) randomly selected sample of physicians or ratings (eg, [14,21,26,36,41]). There is therefore a need for a more comprehensive examination of PRW ratings, to reveal a more generalizable view of ratings and allow trends in rating habits to be identified. As far as we are aware, only 2 studies from the United States [13] and Canada [27] have presented such findings.

This study takes a different approach from most previous studies and analyzes an entire decade of online ratings for all medical specialties on the German PRW, jameda [14,21,26,42] (Please note that the data are not publicly available but may be provided from the provider of the website for research purposes upon request.). Jameda was founded in 2007 and since 2016, has been a wholly owned subsidiary of Burda Digital GmbH. The commercial website provides users with a categorized search function to find suitable physicians, the ability to make appointments with physicians online, the possibility to have video consultations with physicians, an encyclopedia with information from experts on health topics, and an opportunity to rate physicians on a predefined grading system or leave narrative comments. In Germany, a total of 25 PRWs have been identified [8]; however, previous research has indicated that jameda is the German PRW with the highest public awareness, usage, and number of ratings given [4,14,26].

The aim of this study was to examine how ratings posted on the German PRW jameda have developed over the past decade. In particular, it aimed to explore (1) the distribution of ratings according to time-related aspects (year, month, day of the week, and hour of the day) between 2010 and 2019, (2) the number of physicians with ratings, (3) the average number of ratings per physician, (4) the average rating, (5) whether differences exist between medical specialties, and (6) the characteristics of the patients rating physicians.

Methods

Overview

All scaled-survey online ratings that were posted on jameda between 2010 and 2019 were provided by jameda. Ratings on jameda are given according to the 6-point grading system used in German schools (1=very good, 2=good, 3=satisfactory, 4=fair, 5=deficient, and 6=insufficient) [24], in relation to 5 questions: (1) satisfaction with the treatment provided by the physician, (2) the physician's explanation about the illness and treatment, (3) the relationship of trust with the physician, (4) the time the physician spent with the patient, and (5) friendliness of the physician. Additionally, a mean score ("overall performance") is calculated based on the results for Q1 to Q5 [24]. The data also contained the physician's year of birth and medical specialty, as well as the rating patient's gender, age, and health insurance status.

Statistical Analysis

Descriptive statistics included means and SDs for continuous variables as well as numbers and percentages for categorical variables. To analyze whether differences existed between 2 groups, the Mann-Whitney *U* test was used for continuous nonparametric variables, and the Kruskal-Wallis test was applied to determine differences between more than 2 groups. The Shapiro-Wilk test was used to examine the normality of the data distribution. Cohen *d* was calculated to measure the magnitude of the effect size by comparing the standardized difference between the means of 2 groups. All statistical analyses were conducted using SPSS version 22.0 (IBM Corp, Armonk, NY). Differences were considered to be significant if $P<.05$ and highly significant if $P<.001$.

XSL•FO
RenderX

Results

Distribution of Ratings and Mean Ratings

In total, 1,906,146 ratings were posted on jameda between 2010 and 2019 (see Table 1). The highest proportions of ratings were left in 2017 (293,744/1,906,146, 15.41%) and 2018 (292,721/1,906,146, 15.36%). In 2019, there was a decline in the number of ratings (232,739/1,906,146, 12.21%) in comparison with the previous years. Ratings were distributed throughout the months of the year relatively equally (minimum in December: 143,620/1,906,146, 7.53%; maximum in March: 173,865/1,906,146, 9.12%), but more variation was found by day of the week (minimum on Saturdays: 123,024/1,906,146, 6.45%; maximum on Tuesdays: 356,128/1,906,146, 18.68%) and by hour of the day (minimum during 3-4 am: 4659/1,906,146, 0.24%; maximum during 11-12 am: 152,606/1,906,146, 8.00%). Likewise, the mean ratings were relatively similar across years (minimum in 2019: mean 1.71, SD 1.52; maximum in 2013: mean 1.83, SD 1.56), months (minimum in January: mean 1.73, SD 1.49; maximum in August: mean 1.77, SD 1.54), and days (minimum on Sunday: mean 1.68, SD 1.45; maximum on Monday: mean 1.78, SD 1.54). However, more variation could be seen by hour of the day (minimum during 7-8 am: mean 1.67, SD 1.43; maximum during 2-3 am and 3-4 am: mean 2.05, SD 1.75 and mean 2.05, SD 1.72, respectively).

XSL•FO
RenderX

Table 1. Distribution of ratings (N=1,906,146) and mean ratings.

Timeframe	Ratings, n (%)	Mean rating, mean (SD)
Year		
2010	31,908 (1.67)	1.73 (1.42)
2011	61,726 (3.23)	1.74 (1.44)
2012	98,041 (5.14)	1.77 (1.50)
2013	154,119 (8.08)	1.83 (1.56)
2014	219,319 (11.51)	1.81 (1.54)
2015	237,354 (12.45)	1.79 (1.54)
2016	284,475 (14.92)	1.71 (1.48)
2017	293,744 (15.41)	1.73 (1.52)
2018	292,721 (15.36)	1.78 (1.57)
2019	232,739 (12.21)	1.71 (1.52)
Month		
January	170,699 (9.00)	1.73 (1.49)
February	167,728 (8.80)	1.77 (1.53)
March	173,865 (9.11)	1.77 (1.53)
April	151,098 (7.93)	1.77 (1.53)
May	152,995 (8.02)	1.76 (1.53)
June	147,422 (7.73)	1.76 (1.53)
July	160,596 (8.43)	1.77 (1.53)
August	151,544 (7.95)	1.77 (1.54)
September	155,261 (8.15)	1.75 (1.52)
October	161,630 (8.48)	1.77 (1.53)
November	169,688 (8.90)	1.75 (1.52)
December	143,620 (7.53)	1.73 (1.51)
Day of the week		
Monday	342,025 (17.94)	1.78 (1.54)
Tuesday	356,128 (18.68)	1.78 (1.54)
Wednesday	329,457 (17.28)	1.75 (1.52)
Thursday	337,364 (17.70)	1.76 (1.53)
Friday	267,234 (14.02)	1.77 (1.54)
Saturday	123,024 (6.45)	1.74 (1.52)
Sunday	150,914 (7.91)	1.68 (1.45)
Hour of the day		
0-1	23,689 (1.24)	1.96 (1.68)
1-2	11,852 (0.62)	2.00 (1.71)
2-3	6686 (0.35)	2.05 (1.75)
3-4	4659 (0.24)	2.05 (1.72)
4-5	5151 (0.27)	1.98 (1.70)
5-6	9681 (0.51)	1.82 (1.57)
6-7	22,818 (1.20)	1.70 (1.47)
7-8	51,225 (2.69)	1.67 (1.43)
8-9	90,270 (4.74)	1.71 (1.47)

XSL•FO
RenderX

Timeframe	Ratings, n (%)	Mean rating, mean (SD)
9-10	122,461 (6.42)	1.74 (1.50)
10-11	144,834 (7.60)	1.75 (1.51)
11-12	152,606 (8.01)	1.77 (1.53)
12-13	143,618 (7.53)	1.78 (1.54)
13-14	136,245 (7.15)	1.76 (1.53)
14-15	129,596 (6.80)	1.74 (1.50)
15-16	121,427 (6.37)	1.75 (1.52)
16-17	116,451 (6.11)	1.76 (1.53)
17-18	111,075 (5.83)	1.77 (1.54)
18-19	101,968 (5.35)	1.75 (1.53)
19-20	98,494 (5.17)	1.73 (1.52)
20-21	95,222 (5.00)	1.72 (1.51)
21-22	89,447 (4.69)	1.73 (1.51)
22-23	71,515 (3.75)	1.78 (1.54)
23-24	45,156 (2.37)	1.85 (1.60)

Number of Rated Physicians and Ratings Per Rated Physician

Between 2010 and 2019, a total of 127,921 physicians were rated on jameda (see Table 2). The number of rated physicians increased constantly from 19,305 in 2010 to 82,511 in 2018. In 2019, the number of rated physicians decreased to 73,071 rated physicians. The number of ratings that rated physicians received demonstrated an increasing trend. In 2010, 66.94% (12,923/19,305) of all rated physicians were rated only once, 30.88% (5961/19,305) were rated 2-5 times, 1.71% (330/19,305) were rated 6-10 times, and 0.47% (91/19,305) were rated 11-50 times. In 2019, 40.84% (29,843/73,071) of all rated physicians were rated only once, 46.89% (34,262/73,071) were rated 2-5 times, 8.21% (5998/73,071) were rated 6-10 times, 3.93% (2875/73,071) were rated 11-50 times, and 0.13% (93/73,071) were rated more than 50 times. Over the entire decade, 11.43% (14,625/127,921) of all rated physicians were rated once, and 4.23% (5413/127,921) were rated more than 50 times. Please note that the overall numbers cannot be summed up here. For example, one physician received 1 rating in 2010, 3 ratings in 2011, 5 ratings in 2013, 1 rating in 2015, 11 ratings in 2015, 23 ratings in 2017, and 19 ratings in 2019. In sum, this physician was rated 63 times and would be assigned to the category "≥51 Ratings." Similarly, the overall average number of ratings per rated physician increased from 1.65 (SD 1.56) in 2010 to 3.19 (SD 4.69) in 2019. Comparing the number of ratings and rated physicians with the total number of physicians in the German outpatient sector [43], in 2010, 13.64% (19,305/141,461) of all physicians had been rated on jameda, 21.93% (31,335/142,855) in 2011, 29.22% (42,089/144,058) in 2012, 36.36% (53,065/145,933) in 2013, 42.71% (63,182/147,948) in 2014, 45.56% (68,392/150,106) in 2015, 50.51% (76,773/151,989) in 2016, 51.69% (79,799/154,369) in 2017, and 52.46% (82,511/157,288) in 2018 (see also Multimedia Appendix 1). Thus, more than half of all German physicians have been rated online on jameda each year in Germany since 2016.

XSL•FO
RenderX

Table 2. Overall ratings on jameda between 2010 and 2019.

Ratings	Year										Overall (n=127,921)
	2010 (n=19,305)	2011 (n=31,336)	2012 (n=42,089)	2013 (n=53,065)	2014 (n=63,182)	2015 (n=68,392)	2016 (n=76,773)	2017 (n=79,799)	2018 (n=82,511)	2019 (n=73,071)	
Overall number and percentage of rated physicians, n (%)											
1 rating	12,923 (66.94)	18,256 (58.26)	21,133 (50.21)	22,177 (41.79)	22,229 (35.18)	24,512 (35.84)	25,859 (33.68)	26,810 (33.60)	28,971 (35.11)	29,843 (40.84)	14,625 (11.43)
2-5 ratings	5961 (30.88)	11,877 (37.90)	18,389 (43.69)	25,321 (47.71)	31,422 (49.73)	33,751 (49.35)	38,263 (49.84)	39,808 (49.89)	40,602 (49.21)	34,262 (46.89)	31,507 (24.63)
6-10 ratings	330 (1.71)	933 (2.98)	1936 (4.60)	4085 (7.70)	6755 (10.69)	7061 (10.32)	8710 (11.35)	9099 (11.40)	9007 (10.92)	5998 (8.21)	26,285 (20.55)
11-50 ratings	91 (0.47)	259 (0.83)	604 (1.44)	1424 (2.68)	2683 (4.25)	2954 (4.32)	3787 (4.93)	3933 (4.93)	3801 (4.61)	2875 (3.93)	50,091 (39.16)
≥51 ratings	0 (0.00)	11 (0.00)	27 (0.01)	58 (0.11)	93 (0.15)	114 (0.12)	154 (0.20)	149 (0.19)	130 (0.16)	93 (0.13)	5413 (4.23)
Percentage of rated physicians, % (N)	13.64 (141,461)	21.93 (142,855)	29.22 (144,058)	36.36 (145,933)	42.71 (147,948)	45.56 (150,106)	50.51 (151,989)	51.69 (154,369)	52.46 (157,288)	N/A[a]	–[b]
Number of ratings per rated physician											
Mean (SD)	1.65 (1.56)	1.97 (2.51)	2.33 (3.22)	2.90 (4.05)	3.47 (4.84)	3.47 (4.95)	3.71 (5.43)	3.68 (5.09)	3.55 (4.92)	3.19 (4.69)	14.90 (24.04)
Maximum	39	137	151	149	165	154	197	143	215	148	943

[a]N/A: not available.

[b]Not applicable.

Rating Evaluations

Of the 1,906,146 ratings posted between 2010 and 2019, 75.16% (1,432,624/1,906,146) of all ratings were in the best rating category of "very good," and 5.66% (107,912/1,906,146) of the ratings were in the lowest category of "insufficient" (see Table 3). Furthermore, the percentage of ratings on both ends of the rating scale increased over time, from 71.95% (2010) to 78.17% (2019) for very positive ratings and from 3.91% (2010) to 6.12% (2019) for very negative ratings. However, the overall average rating remained relatively constant. The average rating was 1.73 (SD 1.42) in 2010 and 1.71 (SD 1.52) in 2019, with an overall average of 1.76 (SD 1.53).

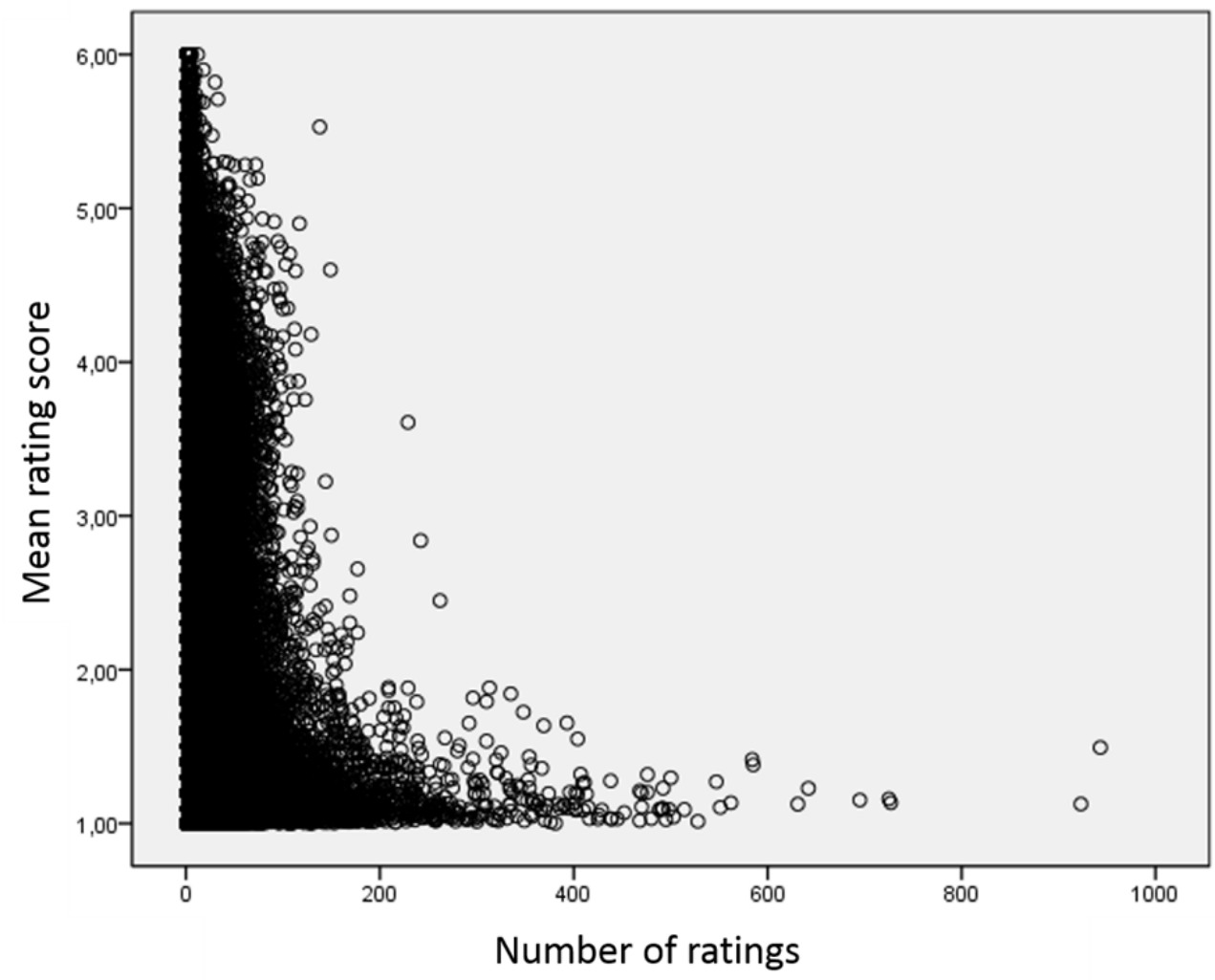
With regards to the correlation between the average rating of a rated physician and the number of ratings per physician, physicians with a lower number of ratings tended to receive ratings across the rating scale, while physicians with a higher number of ratings tended to have better ratings (see Figure 1). Physicians with a single rating had a mean rating of 1.58 (SD 1.28). Afterwards, mean ratings get worse with increasing number of ratings. Physicians with between 21 and 50 online ratings received the worst ratings (mean 1.95, SD 0.84). Mean ratings then improve, with physicians having 51-100 ratings receiving a mean rating of 1.79 (SD 0.86) and physicians with more than 100 ratings receiving the best ratings (mean 1.34, SD 0.47; see Table 4).

XSL•FO
RenderX

Table 3. Overall rating evaluations on jameda between 2010 and 2019.

Overall rating evaluation	Year										Overall (n=1,906,146)
	2010 (n=31,908)	2011 (n=61,726)	2012 (n=98,041)	2013 (n=154,119)	2014 (n=219,319)	2015 (n=237,354)	2016 (n=284,475)	2017 (n=293,744)	2018 (n=292,721)	2019 (n=232,729)	
Rating based on the 6-point grading system, n (%)											
1=very good	22,957 (71.95)	44,952 (72.83)	72,066 (73.51)	111,043 (72.05)	160,263 (73.07)	175,416 (73.90)	217,533 (76.47)	224,527 (76.44)	221,951 (75.82)	181,916 (78.17)	1,432,624 (75.16)
2=good	3406 (10.67)	5783 (9.37)	7889 (8.05)	12,113 (7.86)	16,651 (7.59)	17,328 (7.30)	19,383 (6.81)	18,489 (6.29)	17,205 (5.88)	12,203 (5.24)	130,450 (6.84)
3=satisfactory	1036 (3.25)	2007 (3.25)	2920 (2.98)	4766 (3.09)	6200 (2.83)	6321 (2.66)	6655 (2.34)	6848 (2.33)	6665 (2.28)	4491 (1.93)	47,909 (2.51)
4=fair	1312 (4.11)	2635 (4.27)	4082 (4.16)	6631 (4.30)	9073 (4.14)	9444 (3.98)	9545 (3.36)	9678 (3.29)	10,021 (3.42)	7056 (3.03)	69,477 (3.64)
5=deficient	1948 (6.11)	3910 (6.33)	6233 (6.36)	10,694 (6.94)	15,121 (6.89)	15,658 (6.60)	16,493 (5.80)	17,339 (5.90)	17,537 (5.99)	12,841 (5.52)	117,774 (6.18)
6=insufficient	1249 (3.91)	2439 (3.95)	4851 (4.95)	8872 (5.76)	12,011 (5.48)	13,187 (5.56)	14,866 (5.23)	16,863 (5.74)	19,342 (6.61)	14,232 (6.12)	107,912 (5.66)
Mean (SD)	1.73 (1.42)	1.74 (1.44)	1.77 (1.50)	1.83 (1.56)	1.81 (1.54)	1.79 (1.54)	1.71 (1.48)	1.73 (1.52)	1.78 (1.57)	1.71 (1.52)	1.76 (1.53)

Figure 1. Scatterplot (bivariate) of the number of ratings per physician with the mean overall performance for rated physicians.



XSL•FO
RenderX

Table 4. Online rating results by the number of ratings per physician.

Number of ratings per physician	Average rating, mean (SD)	Rating based on the 6-point grading system[a], n (%)					
		1	2	3	4	5	6
1 (n=14,625)	1.58 (1.28)	11,741 (80.28)	1023 (6.99)	302 (2.06)	394 (2.69)	645 (4.41)	520 (3.56)
2-5 (n=31,505)	1.67 (0.96)	19,733 (62.63)	6204 (19.69)	3477 (11.04)	1395 (4.43)	471 (1.50)	225 (0.71)
6-10 (n=26,258)	1.76 (0.81)	12,505 (47.62)	9459 (36.02)	3096 (11.79)	961 (3.66)	214 (0.81)	23 (0.09)
11-20 (n=29,049)	1.86 (0.78)	12,243 (42.15)	11,188 (38.51)	4289 (14.76)	1162 (4.00)	159 (0.55)	8 (0.02)
21-50 (n=20,658)	1.95 (0.84)	8044 (38.94)	7833 (37.92)	3408 (16.50)	1195 (5.78)	176 (0.85)	2 (0.00)
51-100 (n=3933)	1.79 (0.86)	2084 (52.99)	1122 (28.53)	446 (11.34)	246 (6.25)	35 (0.89)	0 (0.00)
>100 (n=1445)	1.34 (0.47)	1181 (81.73)	220 (15.22)	28 (1.94)	10 (0.69)	5 (0.35)	1 (0.07)
Total (n=127,473)	1.77 (0.92)	67,531 (52.98)	37,049 (29.06)	15,046 (11.80)	5363 (4.21)	1705 (1.34)	779 (0.61)

[a]1=very good, 2=good, 3=satisfactory, 4=fair, 5=deficient, and 6=insufficient.

Ratings by Medical Specialty

Between 2010 and 2019, general practitioners (343,242), internists (266,899), gynecologists (232,914), and orthopedists (229,481) received the highest number of ratings, while pediatricians (87,330), ophthalmologists (79,699), and urologists (63,703) received the lowest number of ratings (see Table 5). However, according to the relative distribution of ratings, the most frequently rated medical specialties in 2018 were orthopedists (6160/7302, 84.36%); oral maxillofacial surgeons (1017/1257, 80.91%); ear, nose, and throat (ENT) specialists (3559/4479, 79.46%); and dermatologists (3562/4632, 76.90%). In contrast, the least frequently rated medical specialties were radiologists (863/4078, 21.16%) and anesthesiologists (601/4247, 14.15%; see Multimedia Appendix 2). Among the 10 most frequently rated medical specialties, the best rated medical specialties were urologists (mean 1.50, SD 1.29), general practitioners (mean 1.64, SD 1.40), and internists (mean 1.68, SD 1.45). The lowest ratings were given to pediatricians (mean 1.92, SD 1.62), ophthalmologists (mean 2.06, SD 1.74), and dermatologists (mean 2.11, SD 1.77).

XSL•FO
RenderX

Table 5. Ratings by medical specialty.

Medical specialty	Year										Overall (n=127,921 rated physicians; n=1,906,146 ratings)
	2010 (n=19,305 rated physicians; n=31,908 ratings)	2011 (n=31,336 rated physicians; n=61,726 ratings)	2012 (n=42,089 rated physicians; n=98,041 ratings)	2013 (n=53,065 rated physicians; n=154,119 ratings)	2014 (n=63,182 rated physicians; n=219,319 ratings)	2015 (n=68,392 rated physicians; n=237,354 ratings)	2016 (n=76,773 rated physicians; n=284,475 ratings)	2017 (n=79,799 rated physicians; n=293,744 ratings)	2018 (n=82,551 rated physicians; n=292,721 ratings)	2019 (n=73,071 rated physicians; n=232,729 ratings)	
General practitioner											
Rated physicians, n (%)	4891 (25.34)	8161 (26.04)	10,533 (25.03)	13,077 (24.64)	15,767 (24.95)	17,016 (24.88)	19,289 (25.12)	19,586 (24.54)	19,967 (24.19)	16,818 (23.02)	33,414 (26.12)
Number of ratings	7241 (22.69)	13,737 (22.25)	19,210 (19.59)	27,952 (18.14)	39,914 (18.20)	42,188 (17.77)	51,504 (18.10)	51,725 (17.61)	51,682 (17.66)	38,089 (16.37)	343,242 (18.01)
Mean rating (SD)[a], n (%)	1.55 (1.21)	1.53 (1.21)	1.54 (1.24)	1.60 (1.33)	1.61 (1.35)	1.66 (1.40)	1.59 (1.35)	1.65 (1.43)	1.73 (1.52)	1.71 (1.51)	1.64 (1.40)
Internist											
Rated physicians, n (%)	3230 (16.73)	5286 (16.87)	6897 (16.39)	8779 (16.54)	10,635 (16.83)	11,511 (16.83)	13,374 (17.42)	13,849 (17.35)	14,634 (17.73)	13,306 (18.21)	23,734 (18.55)
Number of ratings, n (%)	5132 (16.08)	9381 (15.20)	13,697 (13.97)	20,853 (13.53)	29,728 (13.55)	31,611 (13.32)	39,619 (13.93)	40,616 (13.83)	41,642 (14.23)	34,620 (14.88)	266,899 (14.00)
Mean rating (SD)[a]	1.59 (1.27)	1.62 (1.32)	1.62 (1.36)	1.70 (1.44)	1.68 (1.43)	1.70 (1.46)	1.63 (1.40)	1.68 (1.47)	1.73 (1.53)	1.68 (1.49)	1.68 (1.45)
Gynecologist											
Rated physicians, n (%)	2157 (11.17)	3568 (11.39)	5084 (12.08)	6291 (11.86)	7163 (11.34)	7602 (11.12)	8165 (10.64)	8445 (10.58)	8653 (10.48)	7650 (10.47)	11,598 (9.07)
Number of ratings, n (%)	3901 (12.23)	7800 (12.64)	13,987 (14.27)	21,880 (14.20)	28,672 (13.07)	29,795 (12.55)	33,862 (11.90)	34,530 (11.76)	33,562 (11.47)	24,925 (10.71)	232,914 (12.22)
Mean rating (SD)[a]	1.66 (1.36)	1.64 (1.33)	1.69 (1.41)	1.79 (1.49)	1.79 (1.50)	1.80 (1.51)	1.74 (1.49)	1.73 (1.48)	1.76 (1.52)	1.69 (1.47)	1.75 (1.48)
Orthopedist											
Rated physicians, n (%)	1662 (8.61)	2548 (8.13)	3333 (7.92)	4007 (7.55)	4629 (7.33)	5051 (7.39)	5579 (7.27)	5907 (7.40)	6160 (7.46)	5894 (8.07)	8022 (6.27)
Number of ratings, n (%)	3412 (10.69)	6836 (11.07)	11,020 (11.24)	17,805 (11.55)	25,714 (11.72)	28,876 (12.17)	34,242 (12.04)	36,416 (12.40)	35,564 (12.15)	29,596 (12.72)	229,481 (12.04)
Mean rating (SD)[a]	2.08 (1.67)	2.06 (1.67)	2.12 (1.75)	2.15 (1.78)	2.05 (1.72)	1.93 (1.65)	1.82 (1.58)	1.80 (1.57)	1.82 (1.60)	1.70 (1.52)	1.89 (1.63)
Dermatologist (including venereologist)											
Rated physicians, n (%)	855 (4.43)	1354 (4.32)	1947 (4.63)	2467 (4.65)	2811 (4.45)	3003 (4.39)	3229 (4.21)	3415 (4.28)	3562 (4.31)	3232 (4.42)	4517 (3.53)

XSL•FO
RenderX

Medical specialty	Year										Overall (n=127,921 rated physicians; n=1,906,146 ratings)
	2010 (n=19,305 rated physicians; n=31,908 ratings)	2011 (n=31,336 rated physicians; n=61,726 ratings)	2012 (n=42,089 rated physicians; n=98,041 ratings)	2013 (n=53,065 rated physicians; n=154,119 ratings)	2014 (n=63,182 rated physicians; n=219,319 ratings)	2015 (n=68,392 rated physicians; n=237,354 ratings)	2016 (n=76,773 rated physicians; n=284,475 ratings)	2017 (n=79,799 rated physicians; n=293,744 ratings)	2018 (n=82,551 rated physicians; n=292,721 ratings)	2019 (n=73,071 rated physicians; n=232,729 ratings)	
Number of ratings, n (%)	1563 (4.90)	3199 (5.18)	5811 (5.93)	10,461 (6.79)	14,991 (6.84)	15,380 (6.48)	17,513 (6.16)	17,619 (6.00)	17,861 (6.10)	13,355 (5.74)	117,753 (6.18)
Mean rating (SD)[a]	2.06 (1.64)	2.18 (1.71)	2.35 (1.85)	2.28 (1.82)	2.25 (1.82)	2.16 (1.77)	2.05 (1.73)	2.04 (1.74)	2.04 (1.75)	1.94 (1.71)	2.11 (1.77)
ENT[b] specialist, otorhinolaryngologist											
Rated physicians, n (%)	835 (4.33)	1388 (4.43)	1876 (4.46)	2425 (4.57)	2828 (4.45)	3094 (4.52)	3345 (4.36)	3443 (4.31)	3559 (4.31)	3233 (4.42)	4709 (3.68)
Number of ratings, n (%)	1455 (4.56)	3018 (4.89)	5081 (5.18)	9013 (5.85)	13,494 (6.15)	14,626 (6.16)	17,107 (6.01)	16,914 (5.76)	16,118 (5.51)	13,077 (5.62)	109,903 (5.77)
Mean rating (SD)[a]	1.81 (1.50)	1.77 (1.46)	1.76 (1.50)	1.83 (1.57)	1.75 (1.51)	1.74 (1.52)	1.64 (1.43)	1.67 (1.47)	1.75 (1.56)	1.71 (1.53)	1.72 (1.51)
General surgery											
Rated physicians, n (%)	601 (3.11)	1027 (3.28)	1397 (3.32)	1836 (3.46)	2150 (3.40)	2463 (3.60)	2791 (3.64)	3054 (3.83)	3154 (3.82)	2859 (3.91)	4343 (3.40)
Number of ratings, n (%)	1061 (3.33)	2298 (3.72)	3661 (3.73)	6103 (3.96)	9084 (4.14)	10,908 (4.60)	13,240 (4.65)	14,678 (5.00)	14,162 (4.84)	12,272 (5.27)	87,467 (4.59)
Mean rating (SD)[a]	1.80 (1.49)	1.84 (1.57)	1.83 (1.59)	1.83 (1.59)	1.84 (1.61)	1.83 (1.62)	1.79 (1.60)	1.81 (1.63)	1.85 (1.67)	1.78 (1.62)	1.82 (1.62)
Pediatrician											
Rated physicians, n (%)	976 (5.06)	1570 (5.01)	2321 (5.51)	2996 (5.65)	3574 (5.66)	3891 (5.69)	4230 (5.51)	4315 (5.41)	4364 (5.29)	3620 (4.95)	6555 (5.12)
Number of ratings, n (%)	1529 (4.81)	2795 (4.53)	4941 (5.04)	7831 (5.08)	11,059 (5.04)	11,550 (4.87)	13,004 (4.57)	13,295 (4.53)	12,894 (4.40)	8432 (3.62)	87,330 (4.58)
Mean rating (SD)[a]	1.68 (1.35)	1.70 (1.38)	1.76 (1.46)	1.88 (1.57)	1.94 (1.61)	1.94 (1.61)	1.90 (1.60)	1.93 (1.64)	2.03 (1.72)	2.01 (1.73)	1.92 (1.62)
Ophthalmologist											
Rated physicians, n (%)	722 (3.74)	1225 (3.91)	1772 (4.21)	2366 (4.46)	2922 (4.62)	3131 (4.58)	3528 (4.60)	3809 (4.77)	3916 (4.74)	3520 (4.82)	5935 (4.64)
Number of ratings, n (%)	1085 (3.40)	2079 (3.37)	3570 (3.64)	6173 (4.01)	9154 (4.17)	9754 (4.11)	11,899 (4.18)	12,816 (4.36)	12,887 (4.40)	10,282 (4.41)	79,699 (4.18)
Mean rating (SD)[a]	2.07 (1.63)	2.09 (1.67)	2.26 (1.81)	2.20 (1.79)	2.15 (1.78)	2.11 (1.76)	1.98 (1.69)	1.97 (1.69)	2.05 (1.77)	1.96 (1.71)	2.06 (1.74)

XSL•FO
RenderX

Medical specialty	Year										Overall (n=127,921 rated physicians; n=1,906,146 ratings)
	2010 (n=19,305 rated physicians; n=31,908 ratings)	2011 (n=31,336 rated physicians; n=61,726 ratings)	2012 (n=42,089 rated physicians; n=98,041 ratings)	2013 (n=53,065 rated physicians; n=154,119 ratings)	2014 (n=63,182 rated physicians; n=219,319 ratings)	2015 (n=68,392 rated physicians; n=237,354 ratings)	2016 (n=76,773 rated physicians; n=284,475 ratings)	2017 (n=79,799 rated physicians; n=293,744 ratings)	2018 (n=82,551 rated physicians; n=292,721 ratings)	2019 (n=73,071 rated physicians; n=232,729 ratings)	
Urologist											
Rated physicians, n (%)	536 (2.78)	830 (2.65)	1221 (2.90)	1511 (2.85)	1820 (2.88)	1914 (2.80)	2139 (2.79)	2301 (2.88)	2415 (2.93)	2140 (2.93)	3329 (2.60)
Number of ratings, n (%)	845 (2.65)	1639 (2.66)	3221 (3.29)	5141 (3.34)	7207 (3.29)	7753 (3.27)	9556 (3.36)	10,264 (3.49)	9612 (3.28)	8465 (3.64)	63,703 (3.34)
Mean rating (SD)[a]	1.82 (1.54)	1.66 (1.37)	1.57 (1.33)	1.64 (1.41)	1.54 (1.30)	1.50 (1.28)	1.43 (1.20)	1.47 (1.26)	1.49 (1.29)	1.45 (1.25)	1.50 (1.29)
Others											
Rated physicians, n (%)	2840 (14.71)	4379 (13.97)	5708 (13.56)	7310 (13.78)	8883 (14.06)	9716 (14.21)	11,104 (14.46)	11,675 (14.63)	12,127 (14.69)	10,799 (14.78)	21,765 (17.01)
Number of ratings, n (%)	4684 (14.68)	8944 (14.49)	13,842 (14.12)	20,907 (13.57)	30,302 (13.82)	34,913 (14.71)	42,929 (15.09)	44,871 (15.28)	46,737 (15.97)	39,626 (17.03)	287,755 (15.10)
Mean rating (SD)[a]	1.78 (1.48)	1.77 (1.49)	1.68 (1.45)	1.76 (1.54)	1.70 (1.49)	1.65 (1.45)	1.59 (1.40)	1.62 (1.43)	1.66 (1.49)	1.57 (1.41)	1.65 (1.45)

[a]On a 6-point scale: 1=very good, 2=good, 3=satisfactory, 4=fair, 5=deficient, and 6=insufficient.

[b]ENT: ear, nose, throat.

Characteristics of Raters

The rating patients were mostly female (56.8%), between 30 and 50 years old (42.6%), and covered by Statutory Health Insurance (81.0%; see Table 6). However, there were some significant differences between genders, age groups, and health insurance status. Male patients gave significantly more favorable ratings than female patients (mean rating 1.61, SD 1.32 vs. mean 1.77, SD 1.48; *P*<.001). Older patients also gave significantly better ratings than younger patients (*P*<.001). For example, patients aged 51 years or older left a mean rating of 1.52 (SD 1.22), whereas patients aged 29 years or younger left a mean rating of 1.93 (SD 1.59). Finally, patients covered by private health insurance (mean rating 1.43, SD 1.11) gave significantly more favorable evaluations than did patients covered by statutory health insurance (mean rating 1.75, SD 1.47; *P*<.001). Nevertheless, effect sizes were small for all groups, varying between 0.114 and 0.289.

XSL•FO
RenderX

Table 6. Characteristics of raters.

Characteristic	Number of respondents, n (%)	Rating evaluation, mean (SD)	*P* value	Cohen *d*
Gender (n=1,107,092)				
Male	478,592 (43.23)	1.61 (1.32)	<.001[a]	0.114
Female	628,500 (56.77)	1.77 (1.48)		
Age (years; n=1,063,523)				
≤29	164,807 (15.50)	1.93 (1.59)	<.001[b]	0.117[c]; 0.289[d];
30-50	452,774 (42.57)	1.75 (1.46)		0.171[e]
≥51	445,942 (41.93)	1.52 (1.22)		
Health insurance (n=981,635)				
Statutory health insurance	795,107 (81.00)	1.75 (1.47)	<.001[a]	0.245
Private health insurance	186,528 (19.00)	1.43 (1.11)		

[a]Mann-Whitney *U* test.

[b]Kruskal-Wallis test.

[c]≤29 years vs 30-50 years.

[d]≤29 years vs ≥51 or years.

[e]30-50 years vs ≥51 years.

Discussion

This study is one of the most comprehensive analyses of PRW ratings conducted to date and has resulted in a number of key findings: (1) just under 2 million ratings were posted on jameda between 2010 and 2019; (2) a total of 127,921 physicians were rated; (3) the overall average number of ratings per rated physicians nearly doubled; (4) three-quarters of all ratings were in the best rating category of "very good," and the overall average rating remained relatively constant; (5) general practitioners, internists, gynecologists, and orthopedists were the most frequently rated medical specialties; and (6) the rating patients were mostly female, between 30 and 50 years old, and covered by Statutory Health Insurance.

The findings of this study confirm previous research in Germany that indicated that patient ratings show an increasing trend over the past decade [26]. For example, the percentage of all German physicians that had been rated on jameda increased constantly over time from 13.65% (19,305/141,461) in 2010 to 52.46% (82,511/157,288) in 2018. McLennan et al [26] also previously reported that the proportion of physicians from a sample of 298 randomly selected physicians from Hamburg and Thuringia that had been rated at least once had increased between 2010 (range 3.3%-27.8%) and 2014 (range 16.4%-83.2%). Similarly, the average number of ratings per physician also increased between 2010 (range 1.1-3.1) and 2014 (range 1.2-7.5). However, this study only used a small sample from 2 regions in Germany. Overall, there is little international evidence showing the exact development of online ratings over time, which makes it challenging to compare our numbers with those from other similar studies. To the best of our knowledge, more recent studies providing detailed information on a yearly basis are limited. However, 2 studies from the United States [13] and Canada [27] have presented similar findings. First, in 2012, Gao and colleagues [13] showed an increase in the number of rated physicians on RateMDs in the United States from 2475 in 2005 to 112,024 in 2010. Second, Liu and colleagues [27] analyzed a dataset from RateMDs, which included all physicians in Canada in 2018 and showed an increase in the number of ratings for physicians in Canada from 138 in 2005 to 640,603 in 2013. Nevertheless, it should be noted that this study found a plateau in the total number of ratings between 2017 (293,744) and 2018 (292,721). In 2019, a decrease of around 20% in the total number of ratings was seen in comparison with the previous 2 years. In recent years, jameda has implemented and promoted new features on its website (eg, making appointments, video consultations). This has possibly led to lower marketing efforts for collecting online reviews and may also lead to differences from PRWs not offering these addition services. Future studies should investigate whether this latest development can also be observed for other PRWs in Germany and other countries.

This study only provides information regarding jameda. Previous research has demonstrated much lower numbers of both ratings and rated physicians on other German PRWs [4,26]. For example, McLennan and colleagues [26] reported that between 16.4% and 71.1% (mean 41.4%) of physicians were rated on German PRWs overall, compared with 83.2% on jameda. Another study also showed a higher percentage of rated physicians on jameda (90.2%) compared with other relevant German PRWs (32.4% to 61.2%) [4]. Differences in the number of ratings between PRWs can also be shown in the international setting. For example, Trehan and colleagues [44] analyzed online ratings for 250 hand surgeons from the American Society for Surgery of the Hand member directory from 3 PRWs in the United States (HealthGrades, Vitals, RateMDs). Large differences were reported regarding the average number of ratings (13.4, 8.3, and 1.9, respectively) [44]. Further research is required to confirm that this increase in ratings is also true for other PRWs as well.

Furthermore, the percentages of ratings on both ends of the rating scale have increased. This may suggest that a "bimodal" trend in ratings is emerging on jameda, similar to that seen with

XSL•FO
RenderX

the rating of products on websites like Amazon where "amateur" reviewers usually only leave a review because they either love or hate a product [45]. It would be helpful if future research examines if this trend continues and can be found on other PRWs, particularly as this trend is usually not seen on PRWs [26], despite qualitative research in Germany finding that a very positive or very negative experience in the health care relationship is a crucial precondition for patients to be willing to rate a physician [46].

Seven years after the first study on online patient ratings on jameda [18], general practitioners, internists, and gynecologists still receive the highest number of ratings in absolute terms. This does not seem surprising due to the high number of physicians in those medical specialty areas in Germany. Similar to previous research [18], it could also be shown that urologists, general practitioners, and internists were likely to receive more favorable ratings on jameda. In contrast, ophthalmologists and dermatologists are still likely to receive far less favorable ratings. This is also in line with the comprehensive analysis by Liu and colleagues [27] from Canada. Previous research findings have also reported that generalists are more likely to have better online ratings than specialists [10,13]. Qualitative research conducted in Germany by McLennan et al [46] found that factors concerning the physician-patient relationship to be some of the most important influencing people's willingness to rate their physician on PRWs. It is likely that differences in patients' relationships with physicians in various specialties (eg, duration and frequency of contact and the resulting level of trust) is a key factor for this heterogeneity.

The analysis of such a large number of ratings has also provided a more detailed picture of the association between the number of ratings a physician has and their overall evaluation. Although physicians with only 1 rating tended to have very good ratings (81% of all ratings were in the best rating category), this might potentially be explained, at least in part, by "fake ratings" left by physicians themselves or people connected to the physician. Regardless, it certainly calls into question whether results based on a single rating are meaningful at all [7]. Afterwards, more critical rating results were found. In line with previous studies from Germany [18] and the United States [37], the total performance range was found for physicians with a lower number of ratings. This possibly represents a more realistic picture of patient feedback because the percentage of ratings in the very best rating category declined constantly, and it is also likely that those physicians are not using PRWs as a marketing measure to collect a very high number or ratings [18]. However, in contrast to previous research, physicians who received a higher number of ratings were shown to have better ratings. When there were more than 51 ratings, ratings started to improve again, and physicians with more than 100 ratings received by far the most favorable ratings. It is likely that physician with more than 100 ratings are aware of PRWs and are using them as a marketing tool, potentially specifically asking satisfied patients to leave a (positive) rating on a PRW. However, it is possible that these physicians are simply providing outstanding quality of care, leading to the very favorable ratings on PRWs and, subsequently, more patients choosing to use this physician [18]. Future research should examine which assumption is true [18].

In 2019, Pike et al [37] reported a U-shaped relationship between the number of ratings and the overall rating from the Healthgrades website. A negative relationship between the number of ratings and the overall rating could be seen until physicians achieved 21 ratings; thereafter, a positive relationship was seen. It should be noted that, in contrast to jameda, a lower score on Healthgrades means a worse rating (1=poor; 5=excellent). Although regression analysis on the jameda data did not find a satisfying fit, the study provides further broad-scale evidence on the relationship between the number of ratings and the overall evaluation as discussed earlier in this manuscript.

Limitations

The key limitation of this study is that it analyzed online ratings from only a single German PRW, jameda. Although jameda has shown to be the most frequently used German PRW, there are a total of 25 PRWs in Germany [8], and it is unclear how generalizable the results are to other German PRWs or to other countries. In Germany, it would be particularly helpful for future longitudinal research to examine trends in ratings on PRWs run by public health insurers, as previous research has indicated that these PRWs have been able to quickly establish themselves as some of the most used German PRWs alongside jameda [26]. Another limitation of the study is that it only analyzed publicly available ratings; it is not known how many additional ratings jameda received but did not publish or what efforts jameda made to check whether published ratings are genuine and not fake. Indeed, jameda has often been criticized with regards to the number of fake reviews and its business model that offers physicians paid premium profiles. Recent research has raised concerns that online patient feedback is being inappropriately manipulated by many PRWs and that business models that make PRWs reliant on paying physicians may create financial incentives to suppress negative feedback [47]. Although further work is needed on criteria for determining which feedback is published [47], it is also important to have a comprehensive understanding of the ratings that are being viewed by the public on PRWs.

Conclusion

In conclusion, it can be stated that online ratings have been increasing tremendously over the past decade and seem to have become an essential element for patients to leave feedback on the care they receive. More than half of all physicians have been rated online on jameda each year in Germany since 2016. Indeed, with patients increasingly using the internet in relation to their health care [48], it is likely that online patient feedback will become even more important in the future. With online patient feedback mostly positive, physicians do not have to fear online ratings in general; the commonly expressed concerns regarding PRWs being used for "doctorbashing" or defamation [31] or as "platforms for denunciation" [32] have not proven true. Furthermore, less favorable patient ratings often address important elements of a patient-oriented health care system [1] and can help organizations and professionals identify areas that need to be improved [21].

XSL•FO
RenderX

Acknowledgments

This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth through the funding program Open Access Publishing.

The authors would like to thank Ms Elke Ruppert and the management from jameda for the provision of the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The number and percentage of rated physicians in Germany.
[PNG File , 127 KB - jmir_v23i7e24229_app1.png]

Multimedia Appendix 2
Number and distribution of ratings according to the medical specialty (2018).
[DOCX File , 37 KB - jmir_v23i7e24229_app2.docx]

References

1. Powell J, Atherton H, Williams V, Mazanderani F, Dudhwala F, Woolgar S, et al. Using online patient feedback to improve NHS services: the INQUIRE multimethod study. Health Serv Deliv Res 2019 Oct;7(38):1-150. [doi: 10.3310/hsdr07380] [Medline: 31693328]
2. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality-of-care information? A systematic review. Med Care 2009 Jan;47(1):1-8. [doi: 10.1097/MLR.0b013e3181808bb5] [Medline: 19106724]
3. Picker Institute. Using Patient Feedback. National Health Service (NHS) UK. 2009. URL: http://www.nhssurveys.org/Filestore/documents/QIFull.pdf [accessed 2021-07-19]
4. Emmert M, Meszmer N. [A Decade of Online Physician-Rating Websites in Germany: an Assessment of the Current Level of Development]. Gesundheitswesen 2018 Oct;80(10):851-858 [FREE Full text] [doi: 10.1055/s-0043-114002] [Medline: 28759920]
5. Emmert M, Schlesinger M. Patients' Awareness, Usage and Impact of Hospital Report Cards in the US. Patient 2017 Dec;10(6):729-738 [FREE Full text] [doi: 10.1007/s40271-017-0243-y] [Medline: 28447273]
6. Emmert M, Becker S, Meszmer N, Sander U. [Are Facebook Ratings Associated with Hospital Quality and Patient Satisfaction? A Cross-Sectional Analysis of Obstetrics Care in German Hospitals]. Gesundheitswesen 2020 Jun;82(6):541-547 [FREE Full text] [doi: 10.1055/a-0774-7874] [Medline: 30477038]
7. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 01;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
8. Rothenfluh F, Schulz PJ. Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis. J Med Internet Res 2018 Jun 14;20(6):e212 [FREE Full text] [doi: 10.2196/jmir.9105] [Medline: 29903704]
9. Black E, Thompson L, Saliba H, Dawson K, Black NMP. An analysis of healthcare providers' online ratings. Inform Prim Care 2009 Dec 01;17(4):249-253 [FREE Full text] [doi: 10.14236/jhi.v17i4.744] [Medline: 20359403]
10. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
11. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
12. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. J Gen Intern Med 2012 Jun 4;27(6):685-692 [FREE Full text] [doi: 10.1007/s11606-011-1958-4] [Medline: 22215270]
13. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
14. Strech D, Reimann S. [German language physician rating sites]. Gesundheitswesen 2012 Aug;74(8-9):e61-e67 [FREE Full text] [doi: 10.1055/s-0031-1297254] [Medline: 22189671]
15. Greaves F, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National Health Service. J Med Internet Res 2012 Oct 17;14(5):e146 [FREE Full text] [doi: 10.2196/jmir.2280] [Medline: 23076301]

XSL•FO
RenderX

16. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A Mixed-Methods Analysis of Patient Reviews of Hospital Care in England: Implications for Public Reporting of Health Care Quality Data in the United States. The Joint Commission Journal on Quality and Patient Safety 2013 Jan;39(1):7-AP4. [doi: 10.1016/s1553-7250(13)39003-5]
17. Emmert M, Meier F, Pisch F, Sander U. Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. J Med Internet Res 2013 Aug 28;15(8):e187 [FREE Full text] [doi: 10.2196/jmir.2702] [Medline: 23985220]
18. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 06;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
19. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273 [FREE Full text] [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
20. Emmert M, Gerstner B, Sander U, Wambach V. Eine Bestandsaufnahme von Bewertungen auf Arztbewertungsportalen am Beispiel des Nürnberger Gesundheitsnetzes Qualität und Effizienz (QuE). Gesundh ökon Qual manag 2013 Jul 9;19(04):161-167. [doi: 10.1055/s-0033-1335899]
21. Emmert M, Meier F, Heider AK, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73 [FREE Full text] [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
22. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638 [FREE Full text] [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
23. Emmert M, Halling F, Meier F. Evaluations of dentists on a German physician rating Website: an analysis of the ratings. J Med Internet Res 2015 Jan 12;17(1):e15 [FREE Full text] [doi: 10.2196/jmir.3830] [Medline: 25582914]
24. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do Physicians Respond to Web-Based Patient Ratings? An Analysis of Physicians' Responses to More Than One Million Web-Based Ratings Over a Six-Year Period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
25. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-Based Physician Ratings for California Physicians on Probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
26. McLennan S, Strech D, Reimann S. Developments in the Frequency of Ratings and Evaluation Tendencies: A Review of German Physician Rating Websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
27. Liu JJ, Matelski JJ, Bell CM. Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study. J Med Internet Res 2018 Mar 07;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
28. McLennan S. Quantitative Ratings and Narrative Comments on Swiss Physician Rating Websites: Frequency Analysis. J Med Internet Res 2019 Jul 26;21(7):e13816 [FREE Full text] [doi: 10.2196/13816] [Medline: 31350838]
29. McLennan S. The Content and Nature of Narrative Comments on Swiss Physician Rating Websites: Analysis of 849 Comments. J Med Internet Res 2019 Sep 30;21(9):e14336 [FREE Full text] [doi: 10.2196/14336] [Medline: 31573918]
30. Emmert M, Sander U, Kolb B, Meszmer N. Arztbewertungsportale: Die Kritik der Ärzte. Dtsch Arztebl 2017;114(15):731.
31. Carvel J. Patients to rate and review their GPs on NHS website: Doctors say bulletin board will be 'meaningless popularity contest'. The Guardian. 2008 Dec 29. URL: https://www.theguardian.com/society/2008/dec/30/doctors-rating-website-nhs [accessed 2021-07-19]
32. Arztbewertungsportale: Massie Kritik am AOK-Arzt-Navigator: BÄK Newsletter 2009-004, 2009. Ärztlicher Kreisverband Ebersberg. 2009 Jun 25. URL: https://www.aekv-ebersberg.de/aktuelles/105-baek-newsletter-2009-004-vom-25062009 [accessed 2021-07-19]
33. Bundesgerichtshof lehnt den Anspruch eines Arztes auf Löschung seiner Daten aus einem Ärztebewertungsportal. Bundesgerichtshof. 2014. URL: http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&sid=18e188ccd712bed4e8781314de1c4c63&anz=1&pos=0&nr=68915&linked=pm&Blank=1 [accessed 2021-07-19]
34. Ärzte müssen sich anonym im Internet bewerten lassen. Frankfurter Allgemeine. 2014. URL: https://www.faz.net/aktuell/finanzen/meine-finanzen/geld-ausgeben/nachrichten/aerzteportale-jameda-urteil-az-bgh-vi-zr-358-13-13169241.html [accessed 2021-07-19]
35. Emmert M, Meszmer N, Sander U. Do Health Care Providers Use Online Patient Ratings to Improve the Quality of Care? Results From an Online-Based Cross-Sectional Study. J Med Internet Res 2016 Sep 19;18(9):e254 [FREE Full text] [doi: 10.2196/jmir.5889] [Medline: 27644135]
36. Meszmer N, Jaegers L, Schöffski O, Emmert M. [Do online ratings reflect structural differences in healthcare? The example of German physician-rating websites]. Z Evid Fortbild Qual Gesundhwes 2018 Apr;131-132:73-80 [FREE Full text] [doi: 10.1016/j.zefq.2017.11.007] [Medline: 29331281]
37. Pike CW, Zillioux J, Rapp D. Online Ratings of Urologists: Comprehensive Analysis. J Med Internet Res 2019 Jul 02;21(7):e12436 [FREE Full text] [doi: 10.2196/12436] [Medline: 31267982]

XSL•FO
RenderX

38. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: analysis of 2185 reviews. Am J Orthop (Belle Mead NJ) 2014 Aug;43(8):359-363. [Medline: 25136868]
39. Bovenzi CD, Manges KA, Krein H, Heffelfinger R. Online Ratings of Facial Plastic Surgeons: Worthwhile Additions to Conventional Patient Experience Surveys. Facial Plast Surg Aesthet Med 2021;23(2):78-89 [FREE Full text] [doi: 10.1089/fpsam.2020.0049] [Medline: 32716653]
40. Cloney M, Hopkins B, Shlobin N, Dahdaleh NS. Online Ratings of Neurosurgeons: An Examination of Web Data and its Implications. Neurosurgery 2018 Dec 01;83(6):1143-1152 [FREE Full text] [doi: 10.1093/neuros/nyy064] [Medline: 29618127]
41. Haffey SC, Hopman WM, Leveridge MJ. Physicians' Earnings Do Not Affect Their Online Ratings. Front Public Health 2020;8:300 [FREE Full text] [doi: 10.3389/fpubh.2020.00300] [Medline: 32733838]
42. Jameda. URL: https://www.jameda.de/ [accessed 2021-07-17]
43. Die ärztliche Versorgung in Deutschland. Bundesaerztekammer. 2020. URL: https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/ [accessed 2021-07-19]
44. Trehan SK, DeFrancesco CJ, Nguyen JT, Charalel RA, Daluiski A. Online Patient Ratings of Hand Surgeons. J Hand Surg Am 2016 Jan;41(1):98-103 [FREE Full text] [doi: 10.1016/j.jhsa.2015.10.006] [Medline: 26710742]
45. Gilbert E, Karahalios K. Understanding deja reviewers. 2010 Presented at: ACM conference on Computer supported cooperative work; February 6-10, 2010; Savannah, GA URL: https://doi.org/10.1145/1718918.1718961 [doi: 10.1145/1718918.1718961]
46. McLennan S, Strech D, Kahrass H. Why are so few patients rating their physicians on German physician rating websites? A qualitative study. BMC Health Serv Res 2018 Aug 29;18(1):670 [FREE Full text] [doi: 10.1186/s12913-018-3492-0] [Medline: 30157842]
47. McLennan S. Rejected Online Feedback From a Swiss Physician Rating Website Between 2008 and 2017: Analysis of 2352 Ratings. J Med Internet Res 2020 Aug 03;22(8):e18374 [FREE Full text] [doi: 10.2196/18374] [Medline: 32687479]
48. Eysenbach G. The impact of the Internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]

Abbreviations

ENT: ear, nose, and throat
PRW: physician rating website

Edited by R Kukafka, G Eysenbach; submitted 10.09.20; peer-reviewed by S Bidmon, T Kool, Y Mejova; comments to author 02.11.20; revised version received 21.12.20; accepted 30.06.21; published 26.07.21.

Please cite as:
Emmert M, McLennan S
One Decade of Online Patient Feedback: Longitudinal Analysis of Data From a German Physician Rating Website
J Med Internet Res 2021;23(7):e24229
URL: https://www.jmir.org/2021/7/e24229
doi:10.2196/24229
PMID:34309579

©Martin Emmert, Stuart McLennan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Determinants of Fitness App Usage and Moderating Impacts of Education-, Motivation-, and Gamification-Related App Features on Physical Activity Intentions: Cross-sectional Survey Study

Yanxiang Yang[1], MSc; Joerg Koenigstorfer[1], Prof Dr

Chair of Sport and Health Management, Technical University of Munich, Munich, Germany

Corresponding Author:
Joerg Koenigstorfer, Prof Dr
Chair of Sport and Health Management
Technical University of Munich
Georg-Brauchle-Ring 60/62
Campus D – Uptown Munich
Munich, 80992
Germany
Phone: 49 89 289 24559
Email: joerg.koenigstorfer@tum.de

Abstract

Background: Smartphone fitness apps are considered promising tools for promoting physical activity and health. However, it is unclear which user-perceived factors and app features encourage users to download apps with the intention of being physically active.

Objective: Building on the second version of the Unified Theory of Acceptance and Use of Technology, this study aims to examine the association of the seven determinants of the second version of the Unified Theory of Acceptance and Use of Technology with the app usage intentions of the individuals and their behavioral intentions of being physically active as well as the moderating effects of different smartphone fitness app features (ie, education, motivation, and gamification related) and individual differences (ie, age, gender, and experience) on these intentions.

Methods: Data from 839 US residents who reported having used at least one smartphone fitness app were collected via a web-based survey. A confirmatory factor analysis was performed, and path modeling was used to test the hypotheses and explore the influence of moderators on structural relationships.

Results: The determinants explain 76% of the variance in the behavioral intention to use fitness apps. Habit (β=.42; *P*<.001), performance expectancy (β=.36; *P*<.001), facilitating conditions (β=.15; *P*<.001), price value (β=.13; *P*<.001), and effort expectancy (β=.09; *P*=.04) were positively related to behavioral intention to use fitness apps, whereas social influence and hedonic motivation were nonsignificant predictors. Behavioral intentions to use fitness apps were positively related to intentions of being physically active (β=.12; *P*<.001; R^2=0.02). Education-related app features moderated the association between performance expectancy and habit and app usage intentions; motivation-related features moderated the association of performance expectancy, facilitating conditions, and habit with usage intentions; and gamification-related features moderated the association between hedonic motivation and usage intentions. Age moderated the association between effort expectancy and usage intentions, and gender moderated the association between performance expectancy and habit and usage intentions. User experience was a nonsignificant moderator. Follow-up tests were used to describe the nature of significant interaction effects.

Conclusions: This study identifies the drivers of the use of fitness apps. Smartphone app features should be designed to increase the likelihood of app usage, and hence physical activity, by supporting users in achieving their goals and facilitating habit formation. Target group–specific preferences for education-, motivation-, and gamification-related app features, as well as age and gender differences, should be considered. Performance expectancy had a high predictive power for intended usage for male (vs female) users who appreciated motivation-related features. Thus, apps targeting these user groups should focus on goal achievement–related features (eg, goal setting and monitoring). Future research could examine the mechanisms of these moderation effects and their long-term influence on physical activity.

(J Med Internet Res 2021;23(7):e26063) doi:10.2196/26063

XSL•FO
RenderX

KEYWORDS

smartphone; fitness applications; mHealth; technology acceptance; Unified Theory of Acceptance and Use of Technology 2; physical activity; determinants of app usage; education-related app features; motivation-related app features; gamification-related app features; mobile phone

Introduction

Background

To date, there are 3.8 billion smartphone users worldwide [1], and approximately half of them consider their smartphones as something "they could not live without" [2]. Numerous smartphone apps have been developed to allow users to go beyond basic voice calling and texting to social media, gaming, and managing their health and fitness. In June 2021, 98,406 apps in the Google Play Store and 159,758 apps in the Apple App Store were available to users in the health and fitness category [3,4]. These apps aim to promote physical activity and healthy lifestyles [5,6]. It is important to increase our understanding of the factors that influence users in adopting these apps and subsequent associations with intentions to engage in healthy behaviors—both from the perspective of public health and management (eg, app providers)—because stakeholders in these domains are (or should be) interested in finding ways to promote healthy lifestyles via digitization in general and the use of mobile devices in particular.

The most widely used theoretical frameworks that explain why users adopt or use technology are the technology acceptance model [7] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [8]. The two models focus on the organizational context. In consumer settings, the second version of the UTAUT (ie, UTAUT2) has been developed to explain the acceptance of new technology by individuals [9]. Since the first application of UTAUT2 (studying the acceptance of the mobile internet), it has been used to explain smartphone app adoption and usage [10,11], among other applications. With regard to previous empirical studies on mobile health and fitness apps, important gaps exist in the research. First, previous studies have left out the essential determinants that UTAUT2 incorporates (eg, habit and hedonic motivation). Given the importance of habit [12] and hedonic motivation [13], the sole focus on the four determinants proposed by UTAUT seems insufficient [14,15]. Second, the relationship between the intentions to use fitness apps and to be physically active has not been explored. Assessing the downstream effects of intention to use fitness apps is important, because downloaded but unused apps or apps that are unable to motivate people to become or remain physically active will have fewer health benefits [5,16]. Third, understanding whether different fitness app features moderate the relationships of the UTAUT2 determinants and the behavioral intentions of using the app is lacking. Previous research has categorized app features, such as education-related versus motivation-related features [17], but did not consider their influence on structural relationships that aim to explain app usage intentions and physical activity intentions. Finally, despite the fact that the moderating effects of individual-difference variables (eg, age, gender, and experience) have been theorized and empirically assessed [9], they have largely been neglected in prior research on mobile health and fitness apps [18-21]. However, their relevance was shown in a post hoc meta-analysis, for example, in which age was a significant moderator [22].

This study aims to partially fill these gaps and answer four research questions: (1) What are the relationships between the UTAUT2 determinants and behavioral intentions of individuals to use fitness apps? (2) What is the downstream relationship between the behavioral intentions of using fitness apps and being physically active? (3) Do fitness app features moderate the relationships between the UTAUT2 determinants and the intentions of using fitness apps? (4) Are there individual differences regarding age, gender, and user experience in the relationships between the UTAUT2 determinants and intentions to use fitness apps?

To answer the research questions, we applied and extended the UTAUT2 model in the context of smartphone fitness apps. A sample of 839 individuals was surveyed to test our hypotheses. Path modeling was used to test the hypotheses. In the following, we reviewed the extant literature on determinants of fitness app usage, developed the hypotheses, and presented the methodology of our approach.

Literature Review

Smartphone Fitness Apps

Along with the growing consensus on the health benefits of physical activity [23], a myriad of fitness wearables and smartphone fitness apps have been developed to quantify and promote physical activity. Fitness wearables are "devices that offer training plans, assist with activity tracking, and generally collect and process health-related data" [24], whereas fitness apps refer to "the self-contained programs for smartphones designed for the purpose of getting fit" [25]. This study focused on smartphone fitness apps.

Despite the potential of smartphone fitness apps to deliver cost-effective physical activity and health promotion, their effectiveness has not been sufficiently established [5,16,26,27]. In particular, the effectiveness of fitness apps usage or app-based interventions was modest or short-lived [5,16]. In previous studies, only a limited number of factors considered by researchers have been based on theories or behavior change techniques [16,26,27]. Furthermore, only a small number of fitness apps have undergone rigorous evidence-based evaluations in controlled trials [28]. There are some quality concerns in the reporting of these studies, for example, only a few studies have reported whether fitness apps are based on human behavior change theories [28,29]. Herein, we outline the factors that might predict the behavioral intentions of individuals to use fitness apps (and their downstream effects), building upon theories that have been identified as relevant in the information systems literature, particularly UTAUT2.

XSL•FO
RenderX

Determinants of the Behavioral Intentions of Using Fitness Apps

Venkatesh et al [8] developed the UTAUT by integrating eight theories (ie, technology acceptance model, theory of reasoned action, motivational model, theory of planned behavior, combined technology acceptance model, theory of planned behavior model, model of PC utilization, diffusion of innovation theory, and social cognitive theory). According to UTAUT, performance expectancy, effort expectancy, social influence, and facilitating conditions are the four key determinants of behavioral intentions to use technology. In 2012, three additional factors were identified as part of the UTAUT2, namely hedonic motivation, price value, and habit [9]. In the UTAUT2, the individual-difference factors of age, gender, and experience have been identified as important moderators of the relationships between the seven determinants and behavioral intentions. Hew et al [20] applied the UTAUT2 to examine the factors that affect smartphone app adoption in general, considering the moderators of gender and education. They found that all but two factors (ie, social influence and price value) were significant determinants, with habit exerting the strongest influence. Gender and education were nonsignificant moderators. Most important to this research, previous studies used the UTAUT2 to investigate the determinants of behavioral intentions of using fitness-promoting smartwatches [18] and fitness apps [19,30]. However, none of them considered individual-difference factors as moderators, and none of them considered the effect of app features on the proposed relationships.

Specifically, Beh et al [18] found positive relationships among performance expectancy, effort expectancy, facilitating conditions, and hedonic motivation and behavioral intention to use smartwatches for fitness and health monitoring purposes. The authors postulated that perceived vulnerability to developing chronic diseases and perceived severity of chronic diseases would moderate the effects but found only weak support for their hypotheses. Dhiman et al [19] found that effort expectancy, social influence, price value, and habit were positively related to fitness app adoption intentions. They considered self-efficacy to be a predictor of effort expectancy and innovativeness as a predictor of habit; both relationships were significant. Yuan et al [30] did not consider any mediators and found that performance expectancy, hedonic motivation, price value, and habit were predictors of behavioral intentions to continuously use health and fitness apps; however, effort expectancy, social influence, and facilitating conditions were nonsignificant predictors. These studies have important limitations. First, the downstream effects on intentions of being physically active were not assessed in any of the studies. The linkage of fitness app usage intentions and intentions of being physically active is important, because health benefits can only be realized if intended app usage motivates people to become or remain physically active. Second, none of the studies considered app features to be relevant moderators, despite the fact that previous research showed that app features, such as gamification, might moderate the effects of UTAUT2 determinants on app usage intentions [31], and despite the fact that the consideration of risk perception factors (instead of app features) was largely unsuccessful [18]. Third, only one study assessed the moderating roles of age, gender, and experience. However, the authors did not include these variables in the model because of nonsignificant findings [30]. Thus, important similarities with, and differences to the original UTAUT2 studies regarding the influence of age, gender, and experience remain largely unknown. This study aims to fill these gaps partly.

Building upon UTAUT2, we first propose that the seven UTAUT2 determinants relate positively to individuals' intentions to use fitness apps. Second, we postulate positive downstream relationships with the intention of being physically active. Third, we pose a research question that considers three prominent app features (ie, education, motivation, and gamification related) as moderators of the relationships between the seven UTAUT2 determinants and behavioral intentions of using the app. Finally, we explore the moderating effects of individual differences (ie, age, gender, and experience) on the relationship between the seven UTAUT2 determinants and behavioral intentions to use the app. We have listed the hypotheses in the following sections.

Hypotheses Development

Performance Expectancy

Performance expectancy is defined as the "degree to which using a technology will provide benefits to consumers in performing certain activities" [9]. It was the strongest predictor of behavioral intentions in the original UTAUT study [8] and is a pivotal determinant of new technology acceptance in health care [32,33] and fitness wearables [21,34]. In the context of this study, performance expectancy refers to the degree to which a user believes that using a particular fitness app would help improve their fitness. Previous studies have shown a positive relationship between performance expectancy and intention to use fitness apps [15,30]. As the perception that fitness apps help people reach their fitness-related goals should be of high relevance to users, we propose the following:

Hypothesis 1: performance expectancy is positively related to individuals' behavioral intentions to use fitness apps.

Effort Expectancy

Effort expectancy refers to "the degree of ease associated with consumers' use of technology" [9], similar to the perceived ease of use as described in the technology acceptance model [7]. In this study, effort expectancy assesses the perceived ease of use of fitness apps. The easier the individuals believe the fitness apps are to use, the higher is their intention to use them. Prior studies have revealed a positive relationship between effort expectancy and behavioral intention to use fitness apps [15,19] and fitness wearables [18,34]. As people should be interested in intuitive and easy app usage, we expect the following:

Hypothesis 2: effort expectancy is positively related to behavioral intentions of individuals to use fitness apps.

Social Influence

Social influence is defined as "the extent to which consumers perceive that important others (eg, friends, peers) believe they should use a particular technology" [9]. Social influence plays a particular role when users lack information about their usage

XSL•FO
RenderX

[35]. In the context of fitness apps, previous studies have revealed inconsistent results regarding the effect of social influence on behavioral intentions of using fitness apps. It was a positive predictor of usage intentions of students of a Chinese university [15] and Indian users [19], although it did not predict the intentions of college-aged US residents [30]. Given the positive effect of social influence postulated in the original UTAUT2 [9] and the importance of social support in being physically active [36,37], we assume the following:

Hypothesis 3: social influence is positively related to the behavioral intention of individuals to use fitness apps.

Facilitating Conditions

Facilitating conditions refer to "consumers' perceptions of the resources and support available to perform a behavior" [9]. In the context of this research, it reflects the support from resources (eg, ubiquitous internet connection for smartphones) and the required knowledge (eg, experience of smartphone use) to be able to use fitness apps. The original UTAUT2 study [9], as well as studies considering the acceptance of general apps [20] and fitness wearables [18], showed that facilitating conditions increase acceptance. Thus, we postulate the following:

Hypothesis 4: facilitating conditions relate positively to behavioral intentions of individuals to use fitness apps.

Price Value

Price value is defined as "consumers' cognitive trade-off between the perceived benefits of a technology and the monetary cost of using it" [9]. Individuals expect a higher quality of services when they have to pay more for them [30,38]. In the fitness app context, providers offer three main patterns of pricing: free, paid, or freemium (ie, free base app use but additional features need to be paid for). Even if an app can be used for free, individuals might nevertheless consider other cost aspects, such as personal time costs or psychological costs. Previous studies have found a positive relationship between price value considerations and behavioral intentions to use the mobile internet [9], health care wearables [39], and fitness apps [19,30]. Owing to the fact that a high value for a given price can be assumed to be perceived positively by individuals, we propose the following:

Hypothesis 5: price value relates positively to behavioral intentions of individuals to use fitness apps.

Hedonic Motivation

Hedonic motivation refers to "the fun or pleasure derived from using a technology" [9]. If the intrinsic motivation of an individual is high, they typically have high levels of hedonic motivation [40]. A meta-analysis revealed that 58% (53/91) of the included UTAUT2-related empirical studies included hedonic motivation as a factor, whereas 81% (43/53) of the studies found a positive relationship between hedonic motivation and behavioral intentions to use the technology [13]. Hedonic motivation has a positive effect on the intention to adopt health care wearables [18,21] and fitness apps [30]. Thus, we suggest that if a user has fun using a fitness app, they are more likely to use it. Hypothesis 6 is as follows:

Hypothesis 6: hedonic motivation is positively related to the behavioral intentions of individuals to use fitness apps.

Habit

Habit refers to "the extent to which people tend to perform behavior automatically" and was found to be a positive predictor of behavioral intentions to use the mobile internet [9]. Approximately 35% (23/66) of UTAUT2-related empirical studies utilized habit as a construct [12]. Most importantly, 83% (15/18) of the studies revealed positive associations between habit and intention [12]. In the context of this study, we consider habit to be an important predictor, because smartphones are a central means by which individuals can manage and facilitate their daily lives [2] and because individuals use their smartphone (and potentially fitness apps [19,30]) by habit. We thus propose the following:

Hypothesis 7: habit relates positively with the behavioral intentions of individuals using fitness apps.

Downstream Consequence of Behavioral Intentions of Using Fitness Apps

Fitness apps aim to promote user fitness levels. As it is assumable that people who download these apps are (at least partly) committed to reaching this goal, we postulate that higher intentions to use fitness apps relate positively to the willingness of people to be physically active in the future. The claim can be substantiated by consistency theories, arguing that cognitive consistency fosters updates on the expectancy regarding an outcome or a state (here, to be physically active) [41]. However, to date none of the UTAUT2-based studies have examined the relationship between usage intentions of new technology that aims to promote fitness (or health) and the downstream consequence on behavioral intentions to engage in physical activity–related behaviors. Two recent systematic reviews concluded that the effects of fitness apps on physical activity levels are present but are modest in magnitude [5,16]. Previously formed intentions at the individual level might be explanatory variables for these effects. Thus, hypothesis 8 is stated as follows:

Hypothesis 8: behavioral intentions to use fitness apps relate positively to behavioral intentions of being physically active.

Moderating Effects of Fitness App Features

Smartphone apps have certain features, that is, the set of operational functions that an app can perform (eg, gaming). The essence of fitness app features may be summarized within behavior change techniques (eg, goal setting, monitoring, and acquisition of knowledge) [42]. In addition, various frameworks of features implemented in fitness apps have been proposed. For example, Mollee et al [43] identified user input, textual or numerical overviews, social sharing, and general instructions as the most implemented features of fitness apps. Rabin and Bock [44] suggested that fitness tracking, tracking of progress toward fitness goals, and the integration of features that increase enjoyment (eg, music) are user-desired features. Other studies focused on the social features of fitness apps (eg, sharing or comparing steps and receiving social support) [45], whereas a

XSL•FO
RenderX

review concluded that the evidence of social app features to promote fitness was limited [36].

Conroy et al [17] used an empirical approach to cluster fitness apps in terms of features and used cluster analysis to identify two broad categories, namely, motivation related and education related. Motivation-related app features emphasize the social and self-regulation of fitness (eg, tracking, feedback, social support, goal setting, and reward features). Education-related app features focus on fitness education (eg, instruction, coaching, and learning) [17]. These two clusters do not include gamification-related features, which have become relevant in helping individuals improve their health and fitness [46]. Gamification-related features use game design elements to make the user experience playful and enjoyable [47,48]. In this study, we thus consider gamification-related features besides the motivation- and education-related features of fitness apps.

The literature on apps in general (without a focus on physical activity) has considered app features as moderators of the relationship between acceptance determinants and behavioral intentions of using apps [31,48]. However, it remains unclear whether the UTAUT2 determinants interact with fitness app features to explain the behavioral intentions of using these apps. Such interaction effects might explain the modest effects found in systematic reviews on the effects of fitness apps on physical activity [5,16]. To explore this issue, we formulate the following research question: do fitness app features moderate the relationships between the UTAUT2 determinants and behavioral intentions of using fitness apps?

Moderating Effects of Individual Differences

The moderating effects of age, gender, and experience—individual-difference variables—on the relationships between UTAUT2 determinants and behavioral intentions have been proposed and empirically tested in the original UTAUT2 study [9]. In particular, it was theorized that age moderated the relationships between the seven UTAUT2 determinants and behavioral intentions such that the effects are stronger among young (vs old) users for performance expectancy, effort expectancy, and hedonic motivation but weaker for social influence, facilitating conditions, price value, and habit [8,9]. Gender was postulated to moderate the relationship between the seven UTAUT2 determinants and behavioral intentions such that the effects are stronger among women (vs men) for effort expectancy, social influence, facilitating conditions, and price value but weaker for performance expectancy, hedonic motivation, and habit [8,9]. Experience was postulated to moderate the relationships between five UTAUT2 determinants and behavioral intentions such that the effects are stronger among users in the early (vs late) stage of experience for effort expectancy, social influence, facilitating conditions, and hedonic motivation but weaker for habit [8,9]. Three- and four-way interactions of age, gender, and experience were included in the original UTAUT2 study [9]. Despite the fact that the original studies supported these proposed moderator relationships, previous studies on mobile health and fitness apps applying the UTAUT or UTAUT2 did not fully consider them [14,15,18-21,49]. The moderators have been meta-analyzed and suggested as worthy of study [22] or noted as future work [19].

To fill this research gap, we state the following research question: are there individual differences in the relationships between the UTAUT2 determinants and intentions to use fitness apps?

Methods

Study Design and Procedure

This study applied a cross-sectional web-based survey design, and the results were reported according to the CHERRIES checklist [50]. Using a convenience sampling technique, we recruited 867 Amazon Mechanical Turk workers in March 2020. This sample size was considered sufficient based on a thumb rule [51], as well as similar studies on fitness app acceptance [19,30]. Participants were limited to healthy adults who were aged between 18 and 65 years, owned a smartphone, and had downloaded at least one smartphone fitness app. Participants were also required to be able to read and understand English and be located in the United States (ie, US residents). Participants who met the eligibility criteria were invited to participate in the Amazon Mechanical Turk online survey, delivered via Qualtrics. All participants were informed about the study procedures via detailed instructions at the beginning of the survey (Multimedia Appendix 1), including the purpose, inclusion criteria, and estimated time needed to complete the survey. After the instructions were provided, informed consent was obtained from each participant. The survey consisted of UTAUT2-related questions, questions that assessed the dependent variables as well as mediators and moderators, and demographics of participants, which were collected at the end of the survey. Each participant was compensated with US $1.50 for their participation. Once 28 incomplete surveys were eliminated, data from 839 respondents were retained for analysis.

This study was conducted in accordance with the ethical standards of the university faculty board, which acts as the local ethics committee for studies outside the Faculty of Medicine, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Measures

The UTAUT2 items for the seven determinants and behavioral intentions of using apps were adapted to the context of this study [9]. They were measured on a 7-point rating scale ranging from 1 (strongly disagree) to 7 (strongly agree). The behavioral intentions of being physically active were gauged using two separate measures. First, intentions were measured via an adaptation of the International Physical Activity Questionnaire Short Form [52], which covers a period of 4 weeks in the future. The sum of the values (measured in metabolic equivalent of task [MET] min/week) was calculated according to established data processing guidelines [53]. Second, it was measured using a single question: “To what degree do you want to be physically active in the next four weeks?” (1=not at all; 7=very much) [54]. The individual-difference variables of age and gender were self-reported. Experience was measured with a single item: “When did you download a fitness app for the first time? - () months ago,” as done in the original UTAUT2 study [9].

XSL•FO
RenderX

Participants also rated the features of their most preferred app with importance ratings (1=not important at all; 7=extremely important). Importance ratings were used because apps typically have multiple features and because the features from the perspective of users are important in this study [55]. The items for education- and motivation-related app features were formulated in agreement with previous cluster classifications [17] and substantive content of behavior change techniques [42]. Gamification-related app features were operationalized based on the extant literature on gamification and fitness apps [47,56]. All three app features were measured using three items each. Examples of items are as follows: "How important to you are app features that motivate you to be physically active?" for motivation-related features; "How important to you are app features that educate yourself about how to exercise best?" for education-related features; and "How important to you are app features to enjoy yourself while exercising?" for gamification-related features.

Statistical Analyses

Normality was evaluated using multivariate skewness and kurtosis [57]. We conducted a confirmatory factor analysis to evaluate the internal reliability, convergent validity, and discriminant validity of the measurement model [58]. For internal reliability, we examined the Cronbach α (>.70) and construct reliability (>0.70). We used the average variance extracted (AVE; AVE>0.50) and factor loadings for convergent validity [59]. Discriminant validity was assessed using the Fornell-Larcker criterion [59] and the heterotrait-monotrait (HTMT) criteria [60]. Various model fit indices were applied, including the normed chi-square statistic (χ^2/df ratio, value<3.0), Tucker-Lewis index (TLI; TLI>0.90), comparative fit index (CFI; CFI>0.90), root mean square error of approximation (RMSEA; RMSEA<0.05), and standardized root mean square residual (SRMR; SRMR<0.05) [58].

Path modeling (maximum likelihood) was used to test the hypotheses. The variables were mean-centered before the analysis, and gender was coded as a dummy variable (0=female; 1=male). For significant interaction effects between the UTAUT2 determinants and app features, follow-up tests were performed to observe how the moderator changes the hypothesized relationships, as recommended by Dawson [61]. Data analyses were performed using R (RStudio) and the *lavaan* package [62]. The level of significance was set at P<.05.

Participants

A total of 839 participants completed the study. The participants were from 49 US states, with a median of 10 participants per state. They were aged, on average, 37 (SD 10.2) years; 48.3% (405/839) were female; and 51.7% (434/839) were male. Participants were experienced in using fitness apps, as on average they had downloaded the app about 30 months ago. Most participants were White (681/839, 81.2%), employed workers (676/839, 80.6%), married (442/839, 52.7%), and single (322/839, 38.4%). About 66.7% (560/839) reported having a bachelor's degree or higher, whereas 33.3% (279/839) held an associate's degree or lower. They were mostly young adults (562/839, 67% aged between 18 and 40 years), and approximately 44.8% (376/839) of them were either overweight or obese. Approximately 76% (638/839) of them had downloaded two or more fitness apps (mean 3.4, SD 2.5). When asked about their preferred fitness app, 14.1% (118/839) stated MyFitnessPal, 13.2% (111/839) stated Fitbit, and 6.2% (52/839) stated Samsung Health (which are among the preferred apps in real-time app rankings under the category of health and fitness in both the Apple App Store and Google Play Store). In total, 159 different apps were mentioned as the preferred apps by the participants. Table 1 shows the sociodemographic characteristics of the participants.

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of participants (N=839).

Variables	Values
Age (years), mean (SD)	37.3 (10.2)
Gender (female), n (%)	405 (48.3)
BMI[a] (kg/m^2)	
Value, mean (SD)	25.3 (6)
Underweight, n (%)	63 (7.5)
Normal, n (%)	400 (47.7)
Overweight, n (%)	237 (28.3)
Obese, n (%)	139 (16.6)
Education levels, n (%)	
High school degree or below	130 (15.5)
Associate degree	149 (17.8)
College bachelor's degree	390 (46.5)
Master's degree	153 (18.2)
PhD	17 (2)
Marital status, n (%)	
Single (never married)	322 (38.4)
Married	442 (52.7)
Divorced	69 (8.2)
Widowed	6 (0.7)
Income (US $; gross per year), n (%)	
≤15,000	89 (10.6)
15,000-24,999	66 (7.9)
25,000-34,999	104 (12.4)
35,000-49,999	189 (22.5)
50,000-64,999	132 (15.7)
65,000-79,999	122 (14.5)
≥80,000	137 (16.3)
Employment, n (%)	
Employed	676 (80.6)
Self-employed	101 (12)
Unemployed	62 (7.4)
Ethnicity, n (%)	
White	681 (81.1)
Black or African American	84 (10)
Asian	46 (5.5)
Other	28 (3.3)

[a]BMI was classified according to the US Centers for Disease Control and Prevention's BMI weight status categories: underweight (below 18.5 kg/m^2); normal or healthy weight (18.5 to 24.9 kg/m^2); overweight (25.0 to 29.9 kg/m^2); and obese (over 30.0 kg/m^2).

XSL•FO
RenderX

Results

Descriptive Statistics and Assumption Tests

Table 2 provides an overview of the descriptive statistics of the variables. The average ratings of the UTAUT2 determinants ranged from 4.26, for social influence, to 6.02, for facilitating conditions. Education-, motivation-, and gamification-related app features were considered important, with the highest ratings for motivation (mean 5.21) compared with gamification- and education-related app features (mean 5 for both). Participant ratings of their behavioral intentions to use fitness apps were above the midpoint of the scale (mean 5.53); intentions of being physically active in the future were very high for both MET values and the ratings on the seven-point rating scale (mean 4589 MET min/week, SD 3137; and mean 6.07, SD 1.05, respectively). All values of skewness and kurtosis were within the suggested criteria (ie, skewness <2 and kurtosis <7 [63]), indicating normality of the univariate distribution.

XSL•FO
RenderX

Table 2. Measurement model: descriptive statistics, reliability, and convergent validity.

Constructs[a] and items	Value, mean (SD)	Skewness[b]	Kurtosis[b]	Reliability		Convergent validity	
				Cronbach α	Composite reliability	Factor loadings	AVE[c]
Performance expectancy				.87	0.87		0.70
I find the [xx][d] app useful in my daily life	5.54 (1.41)	−1.07	0.88			0.84	
Using the [xx] app helps me accomplish things	5.43 (1.38)	−1.02	0.98			0.86	
Using the [xx] app increases my physical activity levels	5.50 (1.35)	−1.05	1.08			0.80	
Effort expectancy				.89	0.89		0.68
Learning how to use the [xx] app is easy to me	6.02 (1.11)	−1.41	2.52			0.84	
My interaction with the [xx] app is clear and understandable	6.01 (1.09)	−1.4	2.62			0.84	
I find the [xx] app easy to use	6.05 (1.09)	−1.48	2.64			0.86	
It is easy for me to become skillful at using the [xx] app	5.90 (1.12)	−1.27	2.13			0.77	
Social influence				.94	0.94		0.83
People who are important to me think that I should use the [xx] app	4.30 (1.70)	−0.26	−0.56			0.87	
People who influence my behavior think that I should use the [xx] app	4.24 (1.73)	−0.25	−0.64			0.92	
People whose opinions that I value prefer that I use the [xx] app	4.23 (1.72)	−0.29	−0.60			0.94	
Facilitating conditions				.77	0.78		0.54
I have the resources necessary to use the [xx] app	6.08 (1.11)	−1.54	3.03			0.83	
I have the knowledge necessary to use the [xx] app	6.18 (1.05)	−1.53	2.87			0.83	
The [xx] app is compatible with other technologies I use	5.80 (1.29)	−1.24	1.61			0.57	
Hedonic motivation				.91	0.91		0.78
Using the [xx] app is fun	5.07 (1.42)	−0.66	0.27			0.93	
Using the [xx] app is enjoyable	5.24 (1.40)	−0.80	0.50			0.91	
Using the [xx] app is very entertaining	4.71 (1.58)	−0.48	−0.32			0.82	
Price value				.90	0.91		0.76
The [xx] app is reasonably priced	6.28 (1.13)	−1.7	2.59			0.81	
The [xx] app is a good value for the money	6.21 (1.14)	−1.5	1.85			0.93	
At the current price, the [xx] app provides a good value	5.23 (1.15)	−1.72	2.98			0.88	
Habit				.80	0.84		0.66
The use of the [xx] app has become a habit to me	5.34 (1.67)	−1.04	0.33			0.54	
I am addicted to using the [xx] app	3.65 (1.96)	0.09	−1.25			0.87	
I must use the [xx] app	3.84 (1.98)	−0.05	−1.24			0.90	
BI[e]				.89	0.89		0.73

XSL•FO
RenderX

Constructs[a] and items	Value, mean (SD)	Skewness[b]	Kurtosis[b]	Reliability		Convergent validity	
				Cronbach α	Composite reliability	Factor loadings	AVE[c]
I intend to continue using the [xx] app in the future	5.77 (1.37)	−1.41	2.02			0.83	
I will always try to use the [xx] app in my daily life	5.22 (1.55)	−0.92	0.37			0.85	
I plan to continue to use the [xx] app frequently	5.61 (1.45)	−1.27	1.46			0.89	
MO[f]				.85	0.85		0.65
How important to you are app features that motivate you to be physically active?	5.13 (1.54)	−0.88	0.31			0.83	
How important are app features that help you to increase your physical activity levels?	5.38 (1.42)	−1.04	0.88			0.82	
How important to you are app features that remind you to be physically active?	5.11 (1.63)	−0.87	0.17			0.77	
ED[g]				.90	0.90		0.74
How important to you are app features that educate yourself about how to exercise best?	5.01 (1.62)	−0.77	−0.11			0.86	
How important to you are app features that tell you how things work when exercising?	4.87 (1.61)	−0.66	−0.30			0.85	
How important to you are app features that help you do the right things when exercising?	5.11 (1.58)	−0.81	0.12			0.87	
GA[h]				.84	0.84		0.63
How important to you are app features to enjoy yourself while exercising?	5.20 (1.55)	−0.88	0.30			0.86	
How important to you are app features that gamify the exercise experience?	4.62 (1.83)	−0.51	−0.74			0.68	
How important to you are app features that make the exercise experience joyful?	5.16 (1.52)	−0.93	0.47			0.88	
PA[i]				N/A[j]	N/A		N/A
Intentions of being physically active during the next 4 weeks (MET[k] min/week)	4589 (3137)	1.13	1.66			1	
Intentions of being physically active during the next 4 weeks (1-7 rating scale)	6.07 (1.05)	−1.17	1.68			1	
EXP[l]							

XSL•FO
RenderX

Constructs[a] and items	Value, mean (SD)	Skewness[b]	Kurtosis[b]	Reliability		Convergent validity	
				Cronbach α	Composite reliability	Factor loadings	AVE[c]
When did you download a fitness app for the first time? (months ago)	30.07 (25.76)	1.39	2.62	N/A	N/A	1	N/A

[a]Model fit was satisfactory: χ^2_{564}=2112.2; χ^2/df=3.8; comparative fit index=0.93; Tucker-Lewis index=0.91; root mean square error of approximation=0.06; and standardized root mean square residual=0.07.

[b]The criteria for skewness (absolute value <2) and kurtosis (absolute value <7) were fulfilled for a sample size greater than 300 (ie, N=839), indicating normality of the univariate distribution [63].

[c]AVE: average variance extracted.

[d][xx] refers to the brand name of the specified fitness app.

[e]BI: behavioral intentions to use the fitness app.

[f]MO: motivation-related app features.

[g]ED: education-related app features.

[h]GA: gamification-related app features.

[i]PA: Intentions of being physically active. The intentions were measured using the International Physical Activity Questionnaire (metabolic equivalent of task min/week) and a single-item 7-point rating scale. The reported measurement model is based on the first measure.

[j]N/A: not applicable.

[k]MET: metabolic equivalent of task.

[l]EXP: user experience with fitness apps.

Measurement Model

The overall model fit using MET minutes per week values for physical activity intentions as the dependent variable was found to be satisfactory (χ^2_{564}=2112.2; χ^2/df=3.8; CFI=0.93; TLI=0.91; RMSEA=0.06; and SRMR=0.07), after excluding one item for facilitating conditions (ie, “I can get help from others when I have difficulties using the [*brand name*] app” with a factor loading of 0.30). The internal reliability, convergent validity, and discriminant validity of the measurement model were evaluated. All Cronbach α and construct reliability values were ≥.77 (ie, above the suggested threshold of 0.70), indicating internal reliability. The AVE and factor loadings were >0.54, in all cases, above the thresholds of 0.50, suggesting convergent validity (Table 2).

Table 3 shows the results of the discriminant validity. First, no cross-loadings were detected among the measurement items. Second, all the square roots of AVE were greater than the relevant interconstruct correlations with two exceptions (ie, performance expectancy: 0.88; and facilitating conditions: 0.87). The HTMT criteria were fulfilled (ie, all HTMT values were ≤0.85) with one exception (performance expectancy: 0.88), but the value is still within the acceptable range between 0.85 and 0.90 [60].

XSL•FO
RenderX

Table 3. Discriminant validity of the measurement model: Fornell-Larcker criterion and heterotrait-monotrait ratio.

Variables	BI[a]	PE[b]	EE[c]	SI[d]	FC[e]	HM[f]	PV[g]	HA[h]	MO[i]	ED[j]	GA[k]	PA[l]	Age	GEN[m]	EXP[n]
BI	*.856* [o]	.879	.646	.414	.623	.604	.473	.795	.423	.218	.241	N/A[p]	N/A	N/A	N/A
PE	.875	*.835*	.651	.464	.594	.694	.405	.747	.635	.368	.378	N/A	N/A	N/A	N/A
EE	.637	.648	*.823*	.181	.785	.435	.614	.341	.321	.147	.179	N/A	N/A	N/A	N/A
SI	.407	.455	.168	*.911*	.135	.536	.057	.616	.366	.366	.375	N/A	N/A	N/A	N/A
FC	.584	.561	.871	.090	*.733*	.394	.678	.281	.278	.146	.178	N/A	N/A	N/A	N/A
HM	.607	.693	.446	.517	.363	*.881*	.254	.650	.515	.458	.571	N/A	N/A	N/A	N/A
PV	.467	.412	.619	.046	.645	.266	*.873*	.181	.199	.077	.097	N/A	N/A	N/A	N/A
HA	.592	.569	.180	.590	.091	.536	.027	*.811*	.470	.316	.366	N/A	N/A	N/A	N/A
MO	.423	.630	.319	.356	.253	.519	.203	.404	*.806*	.683	.712	N/A	N/A	N/A	N/A
ED	.222	.365	.148	.364	.125	.451	.078	.303	.680	*.861*	.632	N/A	N/A	N/A	N/A
GA	.243	.366	.188	.346	.156	.549	.107	.339	.706	.637	*.794*	N/A	N/A	N/A	N/A
PA	.133	.130	.073	.032	.060	.176	.067	.079	.046	.104	.036	*N/A*	N/A	N/A	N/A
Age	.038	.026	.003	−.036	.053	−.034	.084	.001	.055	−.033	−.011	−.035	*N/A*	N/A	N/A
GEN	.019	.064	.118	−.092	.058	−.038	.041	−.016	.157	.063	.096	−.057	.061	*N/A*	N/A
EXP	.095	.043	.159	−.140	.196	.009	.179	−.099	−.040	−.068	−.061	.084	.051	−.011	*N/A*

[a]BI: behavioral intentions to use the fitness app.

[b]PE: performance expectancy.

[c]EE: effort expectancy.

[d]SI: social influence.

[e]FC: facilitating conditions.

[f]HM: hedonic motivation.

[g]PV: price value.

[h]HA: habit.

[i]MO: motivation-related app features.

[j]ED: education-related app features.

[k]GA: gamification-related app features.

[l]PA: intentions of being physically active.

[m]GEN: gender.

[n]EXP: user experience with fitness apps.

[o]Terms in italics along the diagonal are square roots of average variance extracted. Below the diagonal, the lower left metrics test the discriminant validity according to the Fornell-Larcker criterion. Discriminant validity is fulfilled if the square roots of the average variance extracted are larger than the relevant interconstruct correlations. Furthermore, above the diagonal, the upper right metrics refer to the heterotrait-monotrait ratio, where <0.85 or <0.90 indicates good discriminant validity.

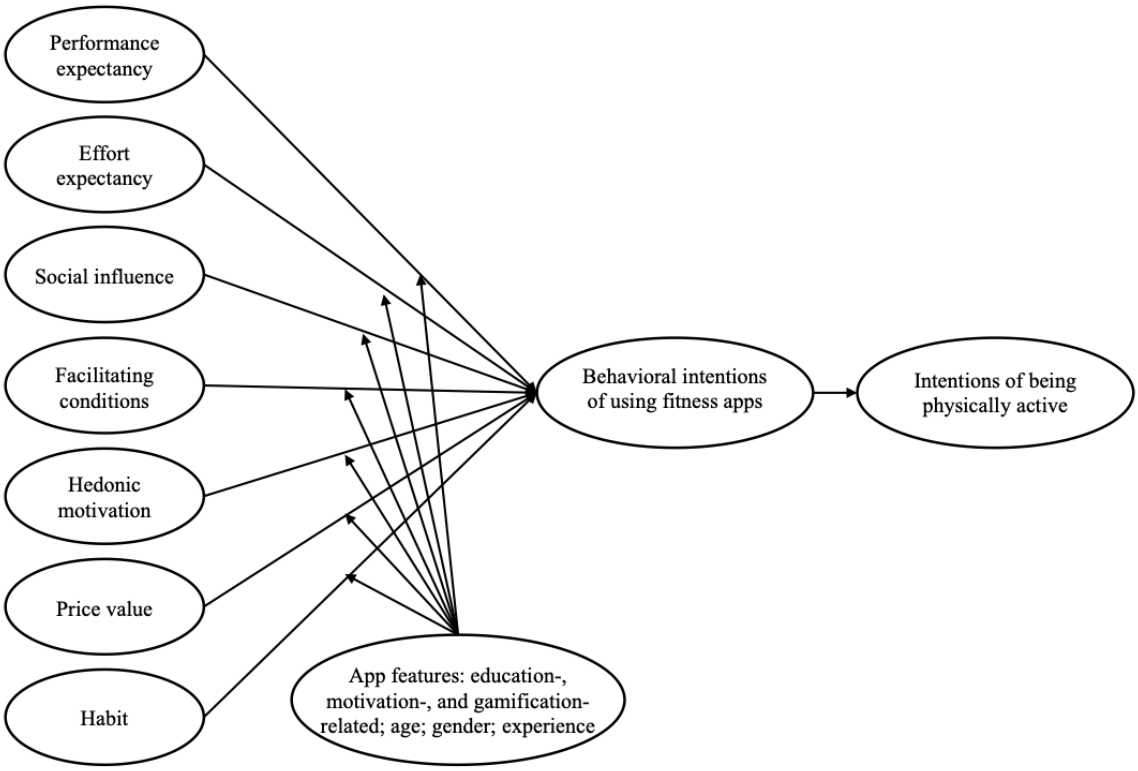
[p]N/A: not applicable.

Structural Model and Hypotheses Testing

Path modeling was used to test the hypotheses. The model was established by modeling the hypothesized paths among the UTAUT2 determinants, behavioral intentions of using fitness apps, intentions of being physically active, and the three app features (Figure 1). On the basis of the different measures of intention to be physically active, two models were established. The first model (considering physical activity intentions measured in MET min/week) had an excellent fit ($\chi^2_{79.00}$=97.74; χ^2/df=1.2; P=.08; CFI=0.984; TLI=0.968; RMSEA=0.017; SRMR=0.006). The model fit for the second model (taking into account physical activity intentions measured on a single-item rating scale) was also good ($\chi^2_{79.00}$=179.07; χ^2/df=2.3; P<.001; CFI=0.925; TLI=0.849; RMSEA=0.039; SRMR=0.010). Both models explained 76% of the variance in the behavioral intentions to use fitness apps.

XSL•FO
RenderX

Figure 1. Hypothesized model for predicting behavioral intentions of using fitness apps and engaging in physical activity based on Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) and the consideration of app features. In agreement with the original UTAUT2 study, experience was postulated to not moderate the relationships between performance expectancy and price value and behavioral intentions of using fitness apps.



In what follows, we first present the results of model 1. Performance expectancy (β=.36, *SE* 0.04; *P*<.001), effort expectancy (β=.09, *SE* 0.04; *P*=.04), facilitating conditions (β=.15, *SE* 0.04; *P*<.001), price value (β=.13, *SE* 0.03; *P*<.001), and habit (β=.42, *SE* 0.04; *P*<.001) were positively related to behavioral intention to use fitness apps, whereas social influence (β=.03, *SE* 0.03; *P*=.37) and hedonic motivation (β=.02, *SE* 0.03; *P*=.63) were nonsignificant predictors. Behavioral intentions to use fitness apps relate positively to intentions of being physically active (β=.12, *SE* 0.03; *P*<.001), explaining 2% of the variance in physical activity intentions. For model 2, the path coefficients between the UTAUT2 determinants and behavioral intentions of using the fitness app were identical to the results obtained from model 1. Behavioral intentions to use fitness apps relate positively to intentions of being physically active (β=.37, *SE* 0.03; *P*<.001), explaining 12% of the variance in physical activity intentions. Thus, hypotheses 1, 2, 4, 6, 7, and 8 were supported, whereas hypotheses 3 and 5 were not supported (Table 4; Figure 2).

XSL•FO
RenderX

Table 4. Path coefficients and hypotheses testing for the seven UTAUT2 determinants and app-feature moderators.

Path	β[a] (*SE*)	Z value	*P* value	Hypothesis testing
UTAUT2[b] determinants				
PE[c]→BI[d]	.36 (0.04)	8.62	<.001	Hypothesis 1 is supported
EE[e]→BI	.09 (0.04)	2.02	.04	Hypothesis 2 is supported
SI[f]→BI	.03 (0.03)	0.90	.37	Hypothesis 3 is not supported
FC[g]→BI	.15 (0.04)	3.55	<.001	Hypothesis 4 is supported
HM[h]→BI	.02 (0.03)	0.49	.63	Hypothesis 5 is not supported
PV[i]→BI	.13 (0.03)	3.97	<.001	Hypothesis 6 is supported
HA[j]→BI	.42 (0.04)	11.52	<.001	Hypothesis 7 is supported
BI→PA[k]	.12 (0.03)	3.60	<.001	Hypothesis 8 is supported
Education-related features				
ED[l]→BI	−.02 (0.03)	−0.89	.37	N/A[m]
ED×PE→BI	−.08 (0.03)	−2.46	.01	N/A
ED×EE→BI	.01 (0.04)	0.17	.86	N/A
ED×FC→BI	.06 (0.04)	1.80	.07	N/A
ED×HM→BI	−.02 (0.03)	−0.76	.45	N/A
ED×PV→BI	−.04 (0.03)	−1.17	.24	N/A
ED×SI→BI	.02 (0.03)	0.70	.48	N/A
ED×HA→BI	.08 (0.03)	2.63	.009	N/A
Motivation-related features				
MO[n]→BI	−.07 (0.03)	−2.34	.02	N/A
MO×PE→BI	.10 (0.03)	3.16	.002	N/A
MO×EE→BI	.08 (0.04)	2.07	.06	N/A
MO×FC→BI	−.11 (0.04)	−2.79	.005	N/A
MO×HM→BI	.02 (0.03)	0.69	.49	N/A
MO×PV→BI	−.03 (0.04)	−0.72	.47	N/A
MO×SI→BI	−.01 (0.03)	−0.47	.64	N/A
MO×HA→BI	−.18 (0.03)	−5.46	<.001	N/A
Gamification-related feature				
GA[o]→BI	−.01 (0.03)	−0.47	.64	N/A
GA×PE→BI	−.03 (0.03)	−0.87	.38	N/A
GA×EE→BI	−.01 (0.04)	−0.29	.77	N/A
GA×FC→BI	−.04 (0.03)	−1.06	.29	N/A
GA×HM→BI	.07 (0.03)	2.77	.006	N/A
GA×PV→BI	.02 (0.03)	0.68	.49	N/A
GA×SI→BI	.01 (0.03)	0.52	.60	N/A
GA×HA→BI	−.04 (0.03)	−1.26	.21	N/A

[a]Unstandardized path coefficient. See Table 5 for the path coefficients of the individual-difference moderators and their interaction effects.

[b]UTAUT2: Unified Theory of Acceptance and Use of Technology 2.

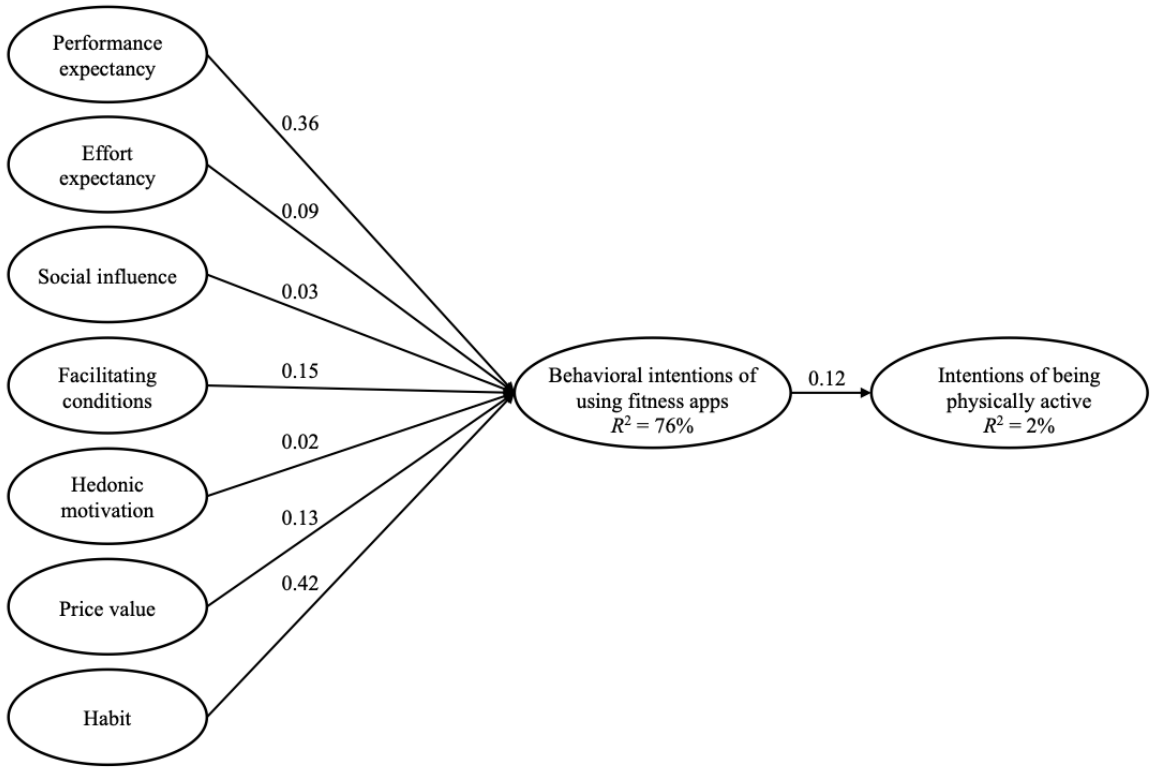
[c]PE: performance expectancy.

[d]BI: behavioral intentions to use the fitness app.

XSL•FO
RenderX

[e]EE: effort expectancy.
[f]SI: social influence.
[g]FC: facilitating conditions.
[h]HM: hedonic motivation.
[i]PV: price value.
[j]HA: habit.
[k]PA: intentions of being physically active, measured in metabolic equivalent of task minutes per week.
[l]ED: education-related app features.
[m]N/A: not applicable.
[n]MO: motivation-related app features.
[o]GA: gamification-related app features.

Figure 2. Path modeling results on the relationship between the Unified Theory of Acceptance and Use of Technology 2 determinants and behavioral intentions of using fitness apps, as well as the downstream effects on intentions of being physically active.



The testing of the interaction effects of app features and the seven UTAUT2 determinants was performed next (Table 4). Education-related app features moderated the relationships between performance expectancy and behavioral intentions to use fitness apps (β=−.08, *SE* 0.03; *P*=.01), as well as between habit and behavioral intentions of using fitness apps (β=.08, *SE* 0.03; *P*=.009). Motivation-related app features moderated the relationships between performance expectancy and behavioral intentions of using fitness apps (β=.10, *SE* 0.03; *P*=.002), facilitating conditions and behavioral intentions to use fitness apps (β=−.11, *SE* 0.04; *P*=.005), and habit and behavioral intentions to use fitness apps (β=−.18, *SE* 0.03; *P*<.001). Gamification-related app features moderated the relationship between hedonic motivation and behavioral intention to use fitness apps (β=.07, *SE* 0.03; *P*=.006).

The testing of the interaction effects of individual differences and the seven UTAUT2 determinants (Table 5) also revealed that age moderated the relationship between effort expectancy and behavioral intention to use fitness apps (β=−.11, *SE* 0.04; *P*=.008). Gender moderated the relationships among performance expectancy and behavioral intention to use fitness apps (β=.13, *SE* 0.06; *P*=.03), habit, and behavioral intentions (β=−.12, *SE* 0.05; *P*=.02). Experience was a nonsignificant moderator. In addition, the joint moderating tests (three- and four-way effects) taking into account individual differences revealed a significant three-way interaction for age, gender, and

XSL•FO
RenderX

hedonic motivation (β=−.14, *SE* 0.06; *P*=.02); a significant three-way interaction for age, experience, and effort expectancy (β=.09, *SE* 0.03; *P*=.007), and a significant three-way interaction of age, experience, and habit on behavioral intentions to use fitness apps (β=−.12, *SE* 0.04; *P*=.004). There were no significant four-way interaction effects.

Subsequently, we conducted follow-up tests to describe how the moderators changed the relationships (Table 6), considering low (−1 SD of the mean) and high (+1 SD of the mean) values of the moderators. First, when education-related features were rated as important, the relationship between performance expectancy and usage intentions was weaker compared with when this feature was rated as unimportant. Second, when education-related features were rated as important, the relationship between habit and usage intentions was stronger compared with when these features were rated as unimportant. Third, when motivation-related features were rated as important, the relationship between performance expectancy and usage intentions was stronger, the relationship between facilitating conditions and usage intentions became nonsignificant, and the relationship between habit and usage intentions was weaker compared with when these features were rated unimportant. Fourth, when gamification-related features were rated as important, the relationship between hedonic motivation and usage intentions was stronger but still nonsignificant compared with when this feature was rated unimportant. Furthermore, the relationship between effort expectancy and usage intentions was positive for younger users but nonsignificant for older users. Finally, the relationship between performance expectancy and usage intentions was stronger among males, whereas the relationship between habit and usage intentions was stronger among females.

XSL•FO
RenderX

Table 5. Path coefficients for the individual-difference moderators and their interaction effects.

Path	β[a] (*SE*)	Z value	*P* value
Age→BI[b]	.03 (0.03)	1.26	.21
Age×PE[c]→BI	.03 (0.04)	0.74	.46
Age×EE[d]→BI	−.11 (0.04)	−2.65	.008
Age×SI[e]→BI	−.04 (0.03)	−1.35	.18
Age×FC[f]→BI	.04 (0.04)	1.08	.28
Age×HM[g]→BI	.02 (0.04)	0.45	.65
Age×PV[h]→BI	.01 (0.03)	0.30	.77
Age×HA[i]→BI	.04 (0.04)	1.05	.29
GEN[j]→BI	.06 (0.04)	1.48	.14
GEN×PE→BI	.13 (0.06)	2.20	.03
GEN×EE→BI	.004 (0.06)	−0.07	.94
GEN×SI→BI	−.04 (0.05)	−0.77	.44
GEN×FC→BI	−.06 (0.06)	−1.03	.30
GEN×HM→BI	.06 (0.05)	1.22	.22
GEN×PV→BI	−.05 (0.05)	−1.01	.31
GEN×HA→BI	−.12 (0.05)	−2.34	.02
EXP[k]→BI	.01 (0.03)	0.55	.58
EXP×EE→BI	−.01 (0.04)	−0.38	.70
EXP×SI→BI	−.02 (0.03)	−0.44	.66
EXP×FC→BI	.05 (0.04)	1.15	.25
EXP×HM→BI	.02 (0.03)	0.75	.46
EXP×HA→BI	.01 (0.03)	0.30	.76
Age×GEN→BI	−.02 (0.04)	−0.55	.58
Age×GEN×PE→BI	.10 (0.06)	1.62	.10
Age×GEN×EE→BI	.04 (0.07)	0.63	.53
Age×GEN×SI→BI	.09 (0.04)	1.96	.052
Age×GEN×FC→BI	−.002 (0.06)	−0.04	.97
Age×GEN×HM→BI	−.14 (0.06)	−2.41	.02
Age×GEN×PV→BI	−.02 (0.05)	−0.32	.75
Age×GEN×HA→BI	−.06 (0.05)	−1.16	.25
EXP×GEN→BI	.06 (0.03)	1.99	.047
EXP×GEN×EE→BI	.10 (0.06)	1.72	.09
EXP×GEN×SI→BI	.06 (0.05)	1.32	.19
EXP×GEN×FC→BI	−.04 (0.06)	−0.62	.54
EXP×GEN×HM→BI	−.07 (0.05)	−1.54	.12
EXP×GEN×HA→BI	−.02 (0.05)	−0.53	.60
Age×EXP→BI	.04 (0.04)	1.10	.27
Age×EXP×EE→BI	.09 (0.03)	2.70	.007
Age×EXP×SI→BI	−.02 (0.03)	−0.45	.65
Age×EXP×FC→BI	−.07 (0.04)	−1.71	.09

XSL•FO
RenderX

Path	β[a] (*SE*)	Z value	*P* value
Age×EXP×HM→BI	.06 (0.04)	1.76	.08
Age×EXP×HA→BI	−.12 (0.04)	−2.85	.004
Age×GEN×EXP→BI	−.002 (0.04)	−0.05	.96
Age×GEN×EXP × EE→BI	−.02 (0.06)	−0.25	.80
Age×GEN×EXP×SI→BI	−.02 (0.05)	−0.41	.70
Age×GEN×EXP×FC→BI	.04 (0.07)	0.58	.56
Age×GEN×EXP×HM→BI	−.09 (0.06)	−1.47	.14
Age×GEN×EXP×HA→BI	.03 (0.05)	0.57	.57

[a]Unstandardized path coefficient. See Table 4 for the path coefficients of the seven UTAUT2 determinants and app-feature moderators.
[b]BI: behavioral intentions to use the fitness app.
[c]PE: performance expectancy.
[d]EE: effort expectancy.
[e]SI: social influence.
[f]FC: facilitating conditions.
[g]HM: hedonic motivation.
[h]PV: price value.
[i]HA: habit.
[j]GEN: gender.
[k]EXP: user experience with fitness apps.

Table 6. Slopes for the relationship of the Unified Theory of Acceptance and Use of Technology 2 determinants with behavioral intentions of using fitness apps at different values of the moderator.

Interactions	Low[a] (−1 SD of mean)			High[b] (+1 SD of mean)		
	Slope	*t* test	*P* value	Slope	*t* test	*P* value
ED[c]×PE[d]	0.36	8.05	<.001	0.28	2.56	.01
ED×HA[e]	0.42	9.39	<.001	0.50	4.56	<.001
MO[f]×PE	0.36	8.05	<.001	0.46	4.20	<.001
MO×FC[g]	0.14	3.13	.002	0.03	0.27	.78
MO×HA	0.42	9.39	<.001	0.24	2.19	.03
GA[h]×HM[i]	0.02	0.45	.66	0.09	0.82	.41
Age×EE[j]	0.09	2.01	.04	−0.02	−0.18	.86
GEN[k]×PE	0.36	8.05	<.001	0.49	4.47	<.001
GEN×HA	0.42	9.39	<.001	0.30	2.74	.006

[a]Low: low moderators.
[b]High: high moderators.
[c]ED: education-related app features.
[d]PE: performance expectancy.
[e]HA: habit.
[f]MO: motivation-related app features.
[g]FC: facilitating conditions.
[h]GA: gamification-related app features.
[i]HM: hedonic motivation.
[j]EE: effort expectancy.
[k]GEN: gender. The results for females (dummy: 0) are reported as low moderators; the results for males (dummy: 1) are reported as high moderators.

XSL•FO
RenderX

Discussion

Principal Findings

The purpose of this study was to examine the influence of the UTAUT2 determinants, as well as the moderating effects of different smartphone fitness app features (ie, education, motivation, and gamification related) and individual differences (ie, age, gender, and experience) on the app usage intentions of individuals and their behavioral intentions of being physically active. The results showed that habit and performance expectancy were the two strongest predictors of intentions of individuals to use fitness apps. The effects of performance expectancy were greater when motivation-related features were rated as important and when education-related features were rated as less important. Moreover, the effects of performance expectancy were greater for males. The effects of habit were greater when education-related features were rated as important and when motivation-related features were rated as less important. Furthermore, the effects of habit were greater for females. Age moderated the relationship between effort expectancy and app usage intention. The intentions of individuals to use fitness apps predicted their intentions of being physically active, using two different means of measuring future physical activity.

Theoretical Contribution

We contribute to the literature on mobile health and physical activity in several ways. Answering the first research question (*What are the relationships between the UTAUT2 determinants and intentions to use smartphone fitness apps?*), we found positive relationships among habit, performance expectancy, facilitating conditions, price value, effort expectancy, and behavioral intentions to use fitness apps. Habit and performance expectancy were found to be the most important predictors of intention to use fitness apps, consistent with prior studies (eg, habit [19,20,30] and performance expectancy [14,15,30]). Positive relationships have also been identified for effort expectancy [18-20], facilitating conditions [18,20,21], and price value [19,21,30].

Social influence was a nonsignificant predictor of intention [18,20,30]. Interestingly, the latter finding is not due to the high domain-specific experience of users (given the nonsignificant interaction effect of social influence and experience), who might have relied less on peer opinions for their evaluations and intentions than low-experience users. Furthermore, in contrast to the original UTAUT2 study [9] and previous studies [18,20,21,30], but in agreement with Dhiman et al [19], we found a nonsignificant relationship between hedonic motivation and app usage intentions. This may be explained by the high demands of fitness app users on app usage to achieve their physical activity goals, compared with the fun or pleasure derived from the apps. However, focusing solely on the four determinants proposed by the first version of UTAUT [14,15,34] may be insufficient. Habit, in particular, is the strongest determinant linked to the intention to use fitness apps in this study.

Answering the second research question (*What is the downstream relationship between the behavioral intentions of using fitness apps and of being physically active?*), we contribute to UTAUT2-based research by showing that app usage intentions have important downstream consequences. In particular, individuals have greater intentions of being physically active when they have higher intentions to use fitness apps. Assessing the downstream effect of intention to use fitness apps is important, because downloaded but unused apps or apps unable to motivate people to become or remain physically active will have little health effects [5,16]. The positive relationship between fitness app usage intentions and physical activity intentions indicates that app usage might motivate people to become or remain active. The findings thus contribute to previous research into whether, and when, mobile health and fitness apps may help individuals become physically active [64,65]. However, it should be noted that the intentions of individuals to be physically active are affected by numerous correlates and determinants (eg, self-efficacy, sociodemographic variables, sport club membership, among others) [66], and the intention-behavior gap is considerable [67]. Thus, adding these factors and incorporating measurements of actual physical activity may be warranted in the future.

Answering the third research question (*Do fitness apps moderate the relationships between the UTAUT2 determinants and intentions of using fitness apps?*), this study contributes to previous research that categorized app features [17] yet ignored their influence on the structural relationships proposed by the UTAUT2. On the basis of our exploratory analysis, we identified six relevant interaction effects. One of the most intuitive findings was that when motivation-related features were rated as important, the relationship between performance expectancy and intentions was strong. Research into goal achievement [68,69] might explain the interaction effect: individuals who are interested in improving their physical activity levels, or keeping them at certain levels, might use the app exactly for this purpose. Among the three features, motivational elements aim most directly to help users stick to their goals and plans [70]; as there is goal congruence, the effect is strong [71]. When motivation-related features were rated as important, the relationship between facilitating conditions and usage intentions was not significant. This makes sense, because people who lack resources and capacities are more dependent on help from others compared with people who do have these resources and capacities, particularly when motivation features are not considered crucial (ie, motivation might “not be the problem”). In addition, when motivation-related features were important, the relationship between habit and intention was weaker compared with when this feature was unimportant. This finding might indicate that when habits have been formed, features that motivate individuals to be active (eg, reminders) become less important to these app users [72].

This study also found that performance expectancy had a greater effect on usage intentions when education-related features were rated as unimportant. In this case, individuals might be less interested in being educated—an aspect that might distract them from achieving their goals. In addition, the effect of habit on usage intention was stronger when education-related features were rated as important. This may be explained by the fact that habits of individuals are formed best when they are exposed to

XSL•FO
RenderX

education-related cues when using an app (eg, how and when to exercise best) [73]. Regarding the interaction between hedonic motivation and gamification-related features, no final conclusions can be drawn. Although research into intrinsic motivation [74] and flow [75] may lead us to propose that intrinsic motivation, as a principal source of enjoyment, may be enhanced by the gamification app features (eg, apps using incommensurate gamification elements [likes]) [76], the follow-up tests did not reach significant levels in this study.

Answering the fourth research question (*Are there individual differences in age, gender, and user experience between the relationships of the UTAUT2 determinants and intentions to use fitness apps?*), we found partly significant, partly nonsignificant moderating effects of age, gender, and experience. First, the relationship between effort expectancy and app usage intentions was stronger among younger individuals, which agrees with the original UTAUT2 study [8,9] and a meta-analysis (ie, age group of those aged 25 to 30 years) [22]. Second, the relationship between performance expectancy and usage intentions was stronger among males, which is consistent with the original UTAUT2 study. In contrast, the relationship between habit and usage intention was stronger among females [9]. Thus, females were not more sensitive to new cues, which might have weakened the effect of habit on behavioral intentions. In the context of fitness apps, females may indeed be prone to cues that help them form health-related habits, because they are interested in health- and body-appearance-related topics. Finally, in this study, experience was a nonsignificant moderator regarding the interaction effects of the UTAUT2 determinants on app usage intentions. Thus, differences in experiences between users might be less relevant today—a time in which smartphone users can easily add and delete new apps and in which users are technology savvy.

Managerial Implications

This study has implications for smartphone app designers and managers. First, they can be advised to focus on habit formation and performance (eg, goal setting) when designing fitness apps and tailoring them to potential users. Meeting users' expectations concerning facilitating conditions, price value, and effort expectancy will also increase the likelihood of the app being accepted. Second, practitioners should highlight certain app features that depend on user preferences. For example, motivation-related features are important drivers of app usage intentions for target group users who value performance (education-related features might be less relevant here); habit formation and facilitating conditions are less important to these individuals. Third, health professionals should consider age and gender differences among users with regard to the effects of effort expectancy (age) as well as performance expectancy and habit (gender). Finally, practitioners may also be advised to monitor whether app usage intentions have a positive correlation with intentions of, or even actual, physical activities so that immediate action can be taken when users lose track of their original goals (having already downloaded the app).

Limitations and Outlook

This study has some limitations. First, the generalizability of our findings is limited. We used a nonrepresentative sample of US residents who owned a smartphone and had previously used fitness apps. Future studies may consider inexperienced people with fitness apps to reveal the influence of UTAUT2 determinants on usage intentions at the early- or preadoption stage. Second, given this research design, we did not consider one specific fitness app, but participants stated their preferred app and rated the features of this app. Thus, we considered a variety of apps (which might be beneficial for external validity, given the myriad of apps on the market [3,4]). Researchers might collaborate with certain providers and use real-world app data and objectively measure actual physical activity to validate our findings. Third, we relied on self-reported physical activity intentions using a single measure and the International Physical Activity Questionnaire Short Form. Overreporting is common for the latter (eg, approximately 84% [77]). Finally, future research could look into the mechanisms of moderation effects on individuals' behavioral intentions to use apps, incorporate app features into mobile health interventions accordingly, and evaluate their long-term influence on physical activity levels.

Acknowledgments

The authors would like to thank the participants of this study. The authors did not receive any external funding from this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online instructions to participants.
[DOCX File , 14 KB - jmir_v23i7e26063_app1.docx]

References

1. Number of smartphone users worldwide from 2016 to 2021. Statista. 2020. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2021-03-12]
2. Melumad S, Pham M. The smartphone as a pacifying technology. J Consum Res 2020;47(2):237-255. [doi: 10.1093/jcr/ucaa005]

XSL•FO
RenderX

3. Apple App Store metrics. PocketGamer. 2020. URL: https://www.pocketgamer.biz/metrics/app-store/categories/ [accessed 2021-06-28]
4. Google Play Store statistics. AppBrain. 2020. URL: https://www.appbrain.com/stats/android-market-app-categories [accessed 2021-06-28]
5. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
6. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):21 [FREE Full text] [Medline: 27034992]
7. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
8. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
9. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
10. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Sys 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
11. Duarte P, Pinho JC. A mixed methods UTAUT2-based approach to assess mobile health adoption. J Bus Res 2019 Sep;102:140-150. [doi: 10.1016/j.jbusres.2019.05.022]
12. Tamilmani K, Rana N, Dwivedi Y. Use of 'habit' is not a habit in understanding individual technology adoption: a review of UTAUT2 based empirical studies. In: Smart Working, Living and Organising. Switzerland: Springer; 2018:277-294. [doi: 10.1007/978-3-030-04315-5_19]
13. Tamilmani K, Rana NP, Prakasam N, Dwivedi YK. The battle of Brain vs. Heart: A literature review and meta-analysis of “hedonic motivation” use in UTAUT2. Int J Inf Manage 2019 Jun;46:222-235. [doi: 10.1016/j.ijinfomgt.2019.01.008]
14. Mohadis H, Ali N. Smartphone application for physical activity enhancement at workplace: would office workers actually use it? In: Proceedings of the International Conference on Information and Communication Technology for the Muslim World (ICT4M). 2018 Presented at: International Conference on Information and Communication Technology for the Muslim World (ICT4M); July 23-25, 2018; Kuala Lumpur, Malaysia. [doi: 10.1109/ict4m.2018.00035]
15. Liu D, Maimaitijiang R, Gu J, Zhong S, Zhou M, Wu Z, et al. Using the unified theory of acceptance and use of technology (UTAUT) to investigate the intention to use physical activity apps: cross-sectional survey. JMIR Mhealth Uhealth 2019 Aug 22;7(9):e13127 [FREE Full text] [doi: 10.2196/13127] [Medline: 31507269]
16. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
17. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
18. Beh PK, Ganesan Y, Iranmanesh M, Foroughi B. Using smartwatches for fitness and health monitoring: the UTAUT2 combined with threat appraisal as moderators. Behav Inf Technol 2019 Nov 19;40(3):282-299. [doi: 10.1080/0144929x.2019.1685597]
19. Dhiman N, Arora N, Dogra N, Gupta A. Consumer adoption of smartphone fitness apps: an extended UTAUT2 perspective. J Indian Bus Res 2019 Nov 17;12(3):363-388. [doi: 10.1108/jibr-05-2018-0158]
20. Hew J, Lee V, Ooi K, Wei J. What catalyses mobile apps usage intention: an empirical analysis. Ind Manag Data Syst 2015;115(7):1269-1291. [doi: 10.1108/IMDS-01-2015-0028]
21. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Ind Manag Data Syst 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/IMDS-03-2015-0087]
22. Angosto S, García-Fernández J, Valantine I, Grimaldi-Puyana M. The intention to use fitness and physical activity apps: a systematic review. Sustainability 2020 Aug 17;12(16):6641. [doi: 10.3390/su12166641]
23. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
24. Canhoto AI, Arp S. Exploring the factors that support adoption and sustained use of health and fitness wearables. J Mark Manag 2016 Oct 26;33(1-2):32-60. [doi: 10.1080/0267257X.2016.1234505]
25. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2015 Jun 17;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]
26. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
27. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014;11(1):97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]

XSL•FO
RenderX

28. Carlo AD, Ghomi RH, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
29. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014;14(1):646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
30. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Apr 28;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
31. Peng W, Yuan S, Ma W. Moderating effects of app type on the intention of continued use of mobile apps among college students. Int J Mob Commun 2018;16(6):715-734. [doi: 10.1504/ijmc.2018.095131]
32. Zailani S, Iranmanesh M, Nikbin D, Beng JK. Determinants of RFID adoption in Malaysia's healthcare industry: occupational level as a moderator. J Med Syst 2015 Aug 16;39(8):74. [doi: 10.1007/s10916-015-0263-x] [Medline: 26077128]
33. Iranmanesh M, Zailani S, Nikbin D. RFID continuance usage intention in health care industry. Qual Manag Health Care 2017;26(2):116-123. [doi: 10.1097/QMH.0000000000000134] [Medline: 28375959]
34. Reyes-Mercado P. Adoption of fitness wearables: insights from partial least squares and qualitative comparative analysis. J Syst Inform Tech 2018 Mar 12;20(1):103-127. [doi: 10.1108/jsit-04-2017-0025]
35. Pal D, Funilkul S, Charoenkitkarn N, Kanthamanon P. Internet-of-things and smart homes for elderly healthcare: an end user perspective. IEEE Access 2018;6:10483-10496. [doi: 10.1109/access.2018.2808472]
36. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
37. McNeill LH, Kreuter MW, Subramanian S. Social environment and physical activity: a review of concepts and evidence. Soc Sci Med 2006 Aug;63(4):1011-1022. [doi: 10.1016/j.socscimed.2006.03.012] [Medline: 16650513]
38. Zeithaml VA. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J Mark 2018 Nov 19;52(3):2-22. [doi: 10.1177/002224298805200302]
39. Kim KJ, Shin D. An acceptance model for smart watches: implications for the adoption of future wearable technology. Internet Res 2015 Aug 03;25(4):527-541. [doi: 10.1108/IntR-05-2014-0126]
40. Magni M, Taylor MS, Venkatesh V. 'To play or not to play': a cross-temporal investigation using hedonic and instrumental perspectives to explain user intentions to explore a technology. Int J Hum-Comput St 2010 Sep;68(9):572-588. [doi: 10.1016/j.ijhcs.2010.03.004]
41. Festinger L. A theory of cognitive dissonance. Sci Am 1962 Oct;207(4):93-102. [doi: 10.1038/scientificamerican1062-93] [Medline: 13892642]
42. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
43. Mollee J, Middelweerd A, Kurvers R, Klein M. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
44. Rabin C, Bock B. Desired features of smartphone applications promoting physical activity. Telemed J E Health 2011 Dec;17(10):801-803. [doi: 10.1089/tmj.2011.0055] [Medline: 22010977]
45. Arigo D, Brown MM, Pasko K, Suls J. Social comparison features in physical activity promotion apps: scoping meta-review. J Med Internet Res 2020 Mar 27;22(3):e15642 [FREE Full text] [doi: 10.2196/15642] [Medline: 32217499]
46. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
47. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
48. Hwang KH, Olmsted SM, Nam SH, Chang BH. Factors affecting mobile application usage: exploring the roles of gender, age, and application types from behaviour log data. Int J Mob Commun 2016;14(3):256-272. [doi: 10.1504/ijmc.2016.076285]
49. Wei J, Vinnikova A, Lu L, Xu J. Understanding and predicting the adoption of fitness mobile apps: evidence from China. Health Commun 2021 Jul 10;36(8):950-961. [doi: 10.1080/10410236.2020.1724637] [Medline: 32041437]
50. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
51. Bentler PM, Chou C. Practical issues in structural modeling. Sociol Methods Res 2016 Jun 30;16(1):78-117. [doi: 10.1177/0049124187016001004]
52. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

XSL•FO
RenderX

53. Sjostrom M, Ainsworth B, Bauman A, Bull F, Hamilton-Craig C, Sallis J. Guidelines for data processing analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. CiNii Articles. 2005. URL: https://ci.nii.ac.jp/naid/10030318551/ [accessed 2021-06-17]
54. Biddle S, Soos I, Chatzisarantis N. Predicting physical activity intentions using goal perspectives and self-determination theory approaches. Eur Psychol 1999 Jun;4(2):83-89. [doi: 10.1027//1016-9040.4.4.83]
55. Dallinga J, Janssen M, van der Werf J, Walravens R, Vos S, Deutekom M. Analysis of the features important for the effectiveness of physical activity-related apps for recreational sports: expert panel approach. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e143. [doi: 10.2196/mhealth.9459] [Medline: 29914863]
56. Cotton V, Patel MS. Gamification use and design in popular health and fitness mobile applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
57. Kim Y, Lee JL. Common mistakes in statistical and methodological practices of sport management research. Meas Phys Educ Exerc 2018 Oct 26;23(4):314-324. [doi: 10.1080/1091367x.2018.1537278]
58. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. New York, United States: Pearson; 2009:1-816.
59. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
60. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
61. Dawson JF. Moderation in management research: what, why, when, and how. J Bus Psychol 2013 May 09;29(1):1-19. [doi: 10.1007/s10869-013-9308-7]
62. Rosseel Y. Lavaan: an R package for structural equation modeling and more. J Stat Soft 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
63. Kim H. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013 Feb;38(1):52-54 [FREE Full text] [doi: 10.5395/rde.2013.38.1.52] [Medline: 23495371]
64. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 May 07;37(18):1428-1438 [FREE Full text] [doi: 10.1093/eurheartj/ehv770] [Medline: 26873093]
65. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
66. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]
67. Rhodes RE, de Bruijn G. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013 May 11;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
68. Baretta D, Bondaronek P, Direito A, Steca P. Implementation of the goal-setting components in popular physical activity apps: review and content analysis. Digit Health 2019;5:2055207619862706 [FREE Full text] [doi: 10.1177/2055207619862706] [Medline: 31360535]
69. Biddle S, Wang CK, Kavussanu M, Spray C. Correlates of achievement goal orientations in physical activity: a systematic review of research. Eur J Sport Sci 2003 Dec;3(5):1-20. [doi: 10.1080/17461390300073504]
70. Bryan JF, Locke EA. Goal setting as a means of increasing motivation. J Appl Psychol 1967 Jun;51(3):274-277. [doi: 10.1037/h0024566] [Medline: 6045632]
71. Morf CC, Weir C, Davidov M. Narcissism and intrinsic motivation: the role of goal congruence. J Exp Soc Psychol 2000 Jul;36(4):424-438. [doi: 10.1006/jesp.1999.1421]
72. Stawarz K, Cox A, Blandford A. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul Republic of Korea p. 2653-2662. [doi: 10.1145/2702123.2702230]
73. Gardner B, Rebar A. Habit formation and behavior change. Oxf Res Encycl Psychol 2019 Apr 26:1-27. [doi: 10.1093/acrefore/9780190236557.013.129]
74. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
75. Csikszentmihalyi M, Csikszentmihalyi IS. Flow: the psychology of optimal experience. New York: Harper & Row; 1990:1-416.
76. Feng W, Tu R, Hsieh P. Can gamification increases consumers' engagement in fitness apps? The moderating role of commensurability of the game elements. J Retail Consum Serv 2020 Nov;57:102229. [doi: 10.1016/j.jretconser.2020.102229]
77. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8(1):115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]

XSL•FO
RenderX

Abbreviations

AVE: average variance extracted
CFI: comparative fit index
HTMT: heterotrait-monotrait
MET: metabolic equivalent of task
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by R Kukafka; submitted 26.11.20; peer-reviewed by C Ochoa-Zezzatti, J Offermann-van Heek; comments to author 09.02.21; revised version received 24.03.21; accepted 04.05.21; published 13.07.21.

Please cite as:
Yang Y, Koenigstorfer J
Determinants of Fitness App Usage and Moderating Impacts of Education-, Motivation-, and Gamification-Related App Features on Physical Activity Intentions: Cross-sectional Survey Study
J Med Internet Res 2021;23(7):e26063
URL: https://www.jmir.org/2021/7/e26063
doi:10.2196/26063
PMID:34255656

©Yanxiang Yang, Joerg Koenigstorfer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effect of 24/7, Digital-First, NHS Primary Care on Acute Hospital Spending: Retrospective Observational Analysis

Sam Winward[1], MBChB, BSc; Tejal Patel[1], MBBS; Mazin Al-Saffar[1], MBBS, BSc; Matthew Noble[1], MBBS

Babylon Health, London, United Kingdom

Corresponding Author:
Sam Winward, MBChB, BSc
Babylon Health
1 Knightsbridge Green
London, SW1X 7QA
United Kingdom
Phone: 44 0330 223 1008
Email: sam.winward@babylonhealth.com

Abstract

Background: Digital health has the potential to revolutionize health care by improving accessibility, patient experience, outcomes, productivity, safety, and cost efficiency. In England, the NHS (National Health Service) Long Term Plan promised the right to access digital-first primary care by March 31, 2024. However, there are few global, fully digital-first providers and limited research into their effects on cost from a health system perspective.

Objective: The aim of this study was to evaluate the impact of highly accessible, digital-first primary care on acute hospital spending.

Methods: A retrospective, observational analysis compared acute hospital spending on patients registered to a 24/7, digital-first model of NHS primary care with that on patients registered to all other practices in North West London Collaboration of Clinical Commissioning Groups. Acute hospital spending data per practice were obtained under a freedom of information request. Three versions of NHS techniques designed to fairly allocate funding according to need were used to standardize or "weight" the practice populations; hence, there are 3 results for each year. The weighting adjusted the populations for characteristics that impact health care spending, such as age, sex, and deprivation. The total spending was divided by the number of standardized or weighted patients to give the spending per weighted patient, which was used to compare the 2 groups in the NHS financial years (FY) 2018-2019 (FY18/19) and 2019-2020 (FY19/20). FY18/19 costs were adjusted for inflation, so they were comparable with the values of FY19/20.

Results: The NHS spending on acute hospital care for 2.43 million and 2.54 million people (FY18/19 and FY19/20) across 358 practices and 49 primary care networks was £1.6 billion and £1.65 billion (a currency exchange rate of £1=US $1.38 is applicable), respectively. The spending on acute care per weighted patient for Babylon GP at Hand members was 12%, 31%, and 54% (£93, P=.047; £223, P<.001; and £389, P<.001) lower than the regional average in FY18/19 for the 3 weighting methodologies used. In FY19/20, it was 15%, 35%, and 51% (£114, P=.006; £246, P<.001; and £362, P<.001) lower. This amounted to lower costs for the Babylon GP at Hand population of £1.37, £4.40 million, and £11.6 million, respectively, in FY18/19; and £3.26 million, £9.54 million, and £18.8 million, respectively, in FY19/20.

Conclusions: Patients with access to 24/7, digital-first primary care incurred significantly lower acute hospital costs.

(J Med Internet Res 2021;23(7):e24917) doi:10.2196/24917

KEYWORDS

primary health care; family practice; general practice; cost; cost analysis; telemedicine; digital technology; digital health; digital care; virtual care; hospital; retrospective; observational; cohort; finance; economics; health services research

Introduction

Health systems across the world are experiencing rising health care costs as a percentage of gross domestic product (GDP) [1]. As the global population continues to age [2], this is expected to worsen. Despite the associated increase in health outcomes, technological improvement is considered a further core driver of health care cost growth [3]. Resultantly, health systems rigorously assess the cost-benefit of such technologies before

XSL•FO
RenderX

implementation. That same rigor has not been consistently applied to digital health solutions; hence, there is increasing concern that their proliferation is outpacing their monitoring and evaluation [4-7].

The COVID-19 pandemic has catalyzed the adoption and growth of digital technologies [8-11], intensifying the need to understand their impact. In particular, the use of telehealth, defined as the "delivery of healthcare services, where patients and providers are separated by distance" [12], has surged. As the world emerges from the pandemic, systems must begin to plan which technologies will form part of the new norm.

The capability of telehealth is well understood in terms of its potential to increase accessibility and patient satisfaction [13-17], increase efficiency [18,19], and improve clinical outcomes [20-22] while remaining safe [16]. The evidence is less clear on its cost-effectiveness. Several large-scale reviews concluded that the majority of interventions are cost-effective [23-25], while others are uncertain [19,26,27], including a Cochrane systematic review that stated the cost-benefit of telemedicine for a health system is unclear [26]. The contrasting nature of findings is in part related to the variety of evaluation methodologies and cost resources assessed; for example, one review stated that the predominant reason for cost savings was reduced travel costs [19]. Generally, telehealth cost-effectiveness studies are limited to a single clinical specialty or service modality but importantly do not consider the perspective of the health system [19]. This can lead to inconclusive assessments; as Rahimi [28] highlights, health systems are often in a state of disequilibrium, and new services can address unmet demand, which can cause net increases in expenditure.

One way to overcome this is to assess "all-cause" health care spending, as two recent digital health studies have done [29,30]. Both studies examined digital care management solutions that leverage proactive approaches and must be patient-focussed to succeed. As a result, they have higher potential to deliver savings to a health care system than do simpler telehealth-as-a-service solutions. Such offerings are most commonly a blend of telehealth, eHealth, and mobile health solutions but generally address a single condition.

This paper considers the management of the entire health care needs of a whole population through the provision of highly accessible, digital-first primary care in the English NHS (National Health Service). The NHS Long Term Plan states that all patients should have the right to choose fully digital-first primary care by March 31, 2024 [31]. Increased accessibility to primary care has been shown to reduce emergency department (ED) attendances [32], but little is known about the overall financial impact this service model will have on the health care system. The aim of this paper was to evaluate the impact that access to 24/7, highly accessible, digital-first primary care has on acute hospital spending.

Methods

Study Setting

Babylon GP at Hand is an NHS general practice in England, which is free at the point of need and provides full NHS primary care services under the General Medical Services contract to people living or working in London, Birmingham, and surrounding areas [33,34].

It was the first NHS general practice to adopt a fully digital-first model of primary care and has been operating across London in this way since November 2017. This means a member's first and main point of contact is digital, with either a smartphone app or a web browser being used to access a virtual appointment. In-person services are available when required at 6 sites across London and 1 in Birmingham (2020).

Babylon GP at Hand is accessible for members day and night 365 days a year, with 80% of all appointments being digital (SG Winward, MD, unpublished data, April 1, 2019, to March 31, 2020). This is more than 3 times the "core hours" stipulated in the standard primary care contract (8 AM-6:30 PM Monday to Friday) [34]. Appointments are therefore available at a time and place convenient for members, and 40% of all appointments occur outside the contracted core hours. Access to medical advice and information is fast; 67% of virtual appointments are available within 2 hours of booking, and 81% of all appointments (including those in person) occur within 48 hours of booking (SG Winward, MD, unpublished data, April 1, 2019, to March 31, 2020); meanwhile, the national average of appointments occurring the same or next day is 49% [35]. Members also have access to a comprehensive suite of digital self-care technology, which can check symptoms; perform a digital health assessment; and monitor symptoms, observations, activity, and mood through the Babylon app.

All services are free for registered members, who must live or work within 40 minutes of a Babylon GP at Hand clinic to be eligible. If previously registered with another NHS practice, the members can switch to their registered GP practice to Babylon GP at Hand. The list size has grown from 3000 to over 100,000 members since November 2017, and the Babylon GP at Hand became the largest single practice in the United Kingdom in August 2020 [33].

Data Sources

The commissioner spending on acute hospital care was compared between patients registered at Babylon GP at Hand and patients registered at other practices in the North West London Collaboration of Clinical Commissioning Groups for the NHS financial years (FYs) of April 1, 2018 to March 31, 2019 (FY18/19); and April 1, 2019, to March 31, 2020 (FY19/20).

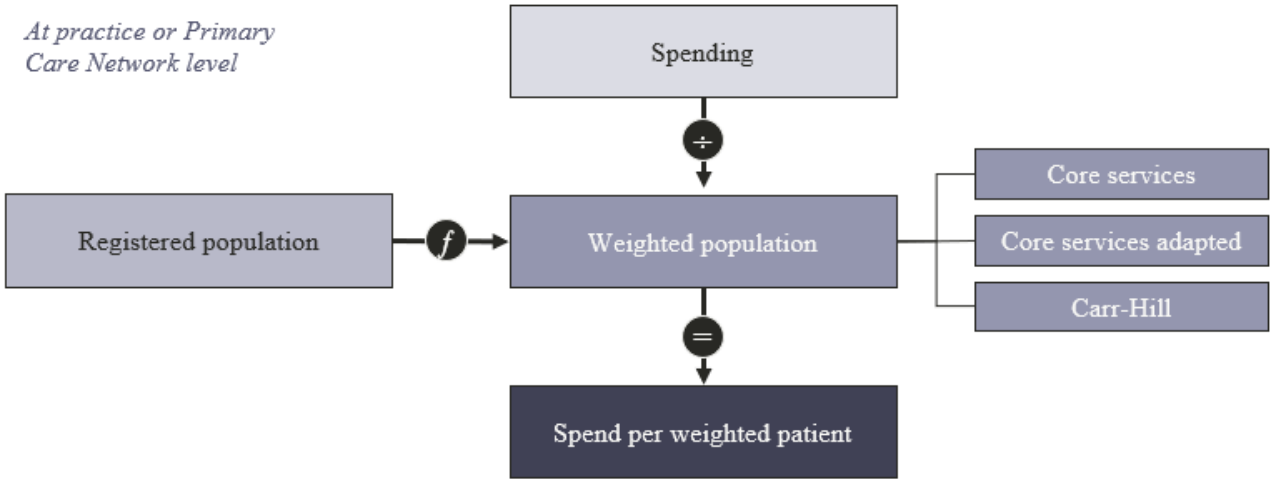
Following a freedom of information (FOI) request, the total acute hospital spending for patients registered at each general practice in North West London Collaboration of Clinical Commissioning Groups was received and aggregated at the level of each practice (Multimedia Appendix 1). The data represented all costs incurred for patients registered at practices in the commissioning region at acute providers. This included acute and general hospital care, such as ED, inpatients, critical care, outpatients, and maternity services, including all associated costs, such as laboratory costs. It did not include mental health spending (eg, inpatient mental health admissions), community spending (eg, district nursing), or specialized services (eg, the

XSL•FO
RenderX

treatment of rare cancers, genetic disorders, or complex medical or surgical conditions), as these services are not consistently provided by all acute trusts or are funded by national rather than local commissioners. Primary care prescription spending was not included in the FOI response, but medications prescribed in hospital were.

Practices were eligible for inclusion if they were active during FY18/19 or FY19/20. Spending was excluded that could not be associated with a patient population.

Figure 1. Overview of methodology.



As the populations registered to each practice were not constant throughout the NHS financial year, an average registered population was calculated based on the practice population at the start of each quarter [33]. All further references to a practice's registered population size are based on this averaged value.

The demographics and health needs of different populations can vary greatly, so each population must be adjusted before their spending can be compared. Existing NHS methodologies were replicated to achieve this. They exist to ensure public funding is distributed fairly by adjusting for characteristics known to impact health care spending, such as, age, sex, and deprivation (Table 1). This transformed the registered population for each practice into a weighted population. For example, if a population with certain characteristics was expected to incur twice the costs of another, it would have twice the number of weighted patients than the other.

Three methods were used to weight the populations, resulting in three weighted populations for each practice (Table 1) [36-39].

Weighting the Populations

Figure 1 shows an overview of the methodology used to compare spending; the total spending per practice is divided by the number of need-adjusted or "weighted" patients. Any differences in the spending per weighted patient between populations would therefore be for reasons other than health care need.

Table 1. Description of 3 methodologies used to weight patient populations to enable the comparison of spending.

Weighting	Description
Core services [36]	Around £80 billion of NHS[a] funding per year is distributed to commissioners for "core services". The core services formula is used to ensure fair distribution of this amount to populations with different characteristics. It has separate components that weight each practice's population for their need for services, including general and acute, mental health, maternity, community care, and prescription needs. In addition, each population is adjusted for local factors: health care utilization, supply of health care services, pricing, unavoidable costs, unmet need (with premature mortality rate used as a proxy), local deprivation, and costs due to unavoidable smallness.
Core services adapted	Three adaptations were made to the core services methodology to better match the spending data received in the freedom of information request. First, as the spending data received did not include mental health, community, or prescription costs, these elements were removed from the formula. Second, the Babylon GP at Hand population was reweighted by the actual age and sex characteristics as opposed to the estimates provided in the core services file. As Babylon GP at Hand is a fast-growing practice, it was over 3 times the size of the estimation [37]. Third, the core services formula assumes that all Babylon GP at Hand patients live in Hammersmith and Fulham, whereas the majority live in other clinical commissioning groups. To improve the accuracy, the local components from each patient's home residence were used.
Carr-Hill [38,39]	The Carr-Hill weighting is used to distribute the global sum, the largest component of primary care funding. It adjusts the population based on drivers of need, including the consulting time recorded for patients with certain characteristics, local premature mortality rates, market forces (local costs), practices rurality index (though this has been phased out), and the number of nursing home patients registered to the practice.

[a]NHS: National Health Service.

XSL•FO
RenderX

Cost Per Weighted Patient

The total spending for each practice was divided by its number of weighted patients, giving the spending per weighted patient for that practice. This was also performed at the level of the primary care networks, which are groups of practices that work collaboratively, totalling around 50,000 people each [40,41]. Primary care networks (groupings of general practices) are more similar in size to the Babylon GP at Hand practice than are other practices and hence are a more appropriate grouping to compare against.

Adjustments for the Babylon GP at Hand Practice

To increase the accuracy of the calculation for Babylon GP at Hand, 3 adjustments were applied.

First, patients registered at Babylon GP at Hand's Birmingham site were removed. As spending data were only received for hospitals in Greater London, no reciprocal adjustments were made to the spending. The Birmingham site opened in June 2019, so this only affected FY19/20.

Second, Babylon GP at Hand launched from an existing practice in July 2017. A cohort of patients who lived near to the Hammersmith and Fulham site continued to receive a traditional model of primary care from the existing provider. This population was removed to better assess the effects of Babylon GP at Hand model of care. As spend data were only received at a practice level, this group of patients was assigned the average weighting and average cost per patient for the Hammersmith and Fulham Clinical Commissioning Group to remove them from the Babylon GP at Hand practice totals.

Third, an independent review of Babylon GP at Hand, commissioned by NHS England, reported that patients who joined the practice were less likely to use certain hospital services in the 12 months prior to joining than were a matched population [42]. If this observation persisted after patient joining, this would imply that the Babylon GP at Hand population would incur lower costs even after adjustment for the characteristics of its populations. To conservatively account for this potential effect, the cost per weighted patient for Babylon GP at Hand was inflated. To determine the degree of inflation, population data, national reference costs, and attendance rates for each age and sex category were used to calculate the expected costs for the Babylon GP at Hand population [43,44]. The same calculation was then performed using the attendance rates modified by the findings in the independent review [42]. A 12% difference between the 2 represented the degree to which the actual cost would be lower than the expected cost, and hence the Babylon GP at Hand spending per weighted patient was increased by this factor.

Adjustment for Inflation

Costs for FY18/19 were adjusted for inflation to be equivalent to FY19/20 values. The GDP deflator at market prices for the United Kingdom was used as produced by Her Majesty's Treasury and published by the Office of National Statistics [45].

Data Analysis

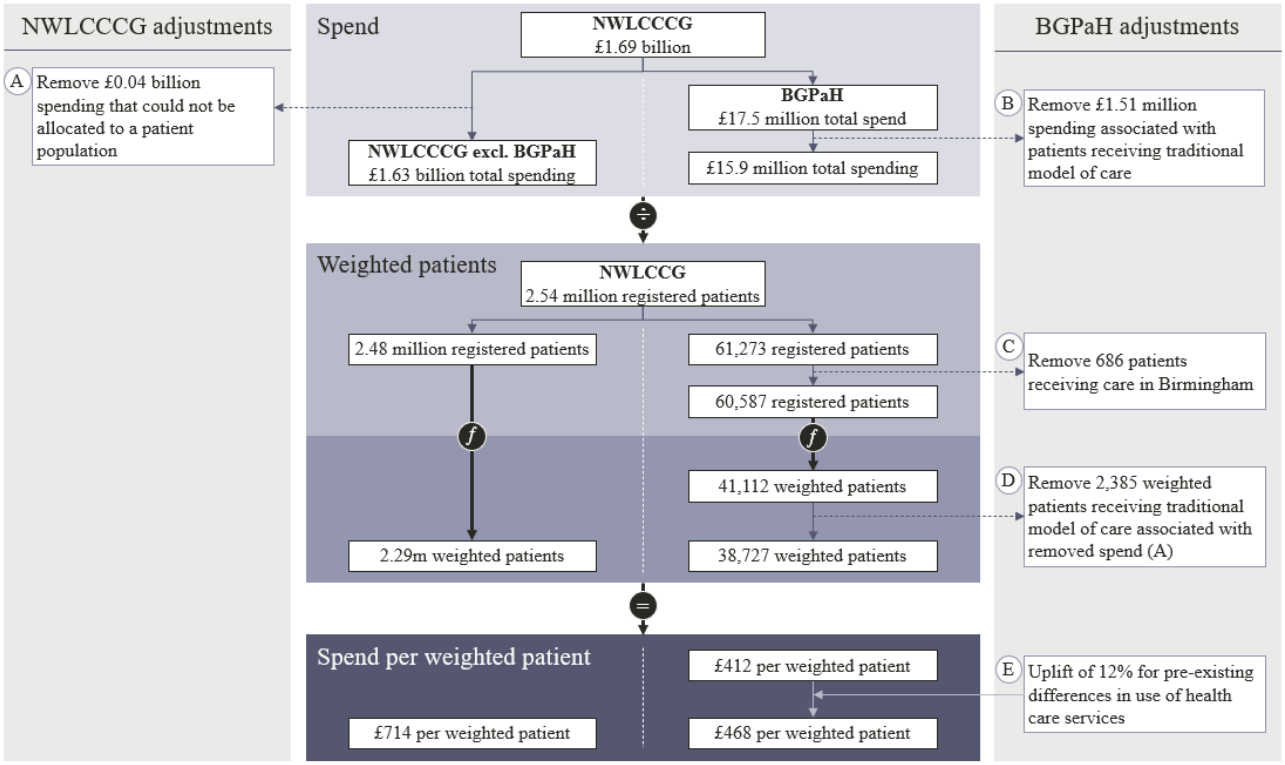
Analysis was performed on SciPy package version 1.5.4 (Python) [46]. Shapiro-Wilk and Kolmogorov-Smirnov tests were used to determine if the data were normally distributed before a 1-sided, simple *z* test was performed to compare the spending per weighted patient between groups.

Results

Figure 2 follows the methodology outlined in Figure 1 to calculate the cost per weighted patient, with the core services methodology being used in FY19/20, including all adjustments.

XSL•FO
RenderX

Figure 2. Flow chart demonstrating how the spending per weighted patient for the core services methodology was calculated for the financial year 2019-2020. BGPaH: Babylon GP at Hand practice; NWLCCCG: North West London Collaboration of Clinical Commissioning Groups.



Spending

The total acute care spending returned in the FOI was £1.64 billion in FY18/19 and £1.69 billion in FY19/20 (a currency exchange rate of £1=US $1.38 is applicable), across 361 practices. An expenditure of £44.6 million (2.71%) in FY18/19 and £43.4 million (2.56%) in FY19/20 was excluded that was associated with 3 practices and an “Unknown Primary Care Network” in the FOI response, as the costs could not be attributed to a patient population (Figure 2A). The remaining spending was £1.6 billion in FY18/19 and £1.65 billion in FY19/20 across 358 practices including Babylon GP at Hand.

The spending for Babylon GP at Hand practice was £8.6 million in FY18/19 and £15.9 million in FY19/20,after the spending associated with patients receiving a traditional model of care (£1.59 million and £1.51 million, respectively) was removed (Figure 2B).

Registered Population

The 358 practices had a total registered population of 2.43 million patients in FY18/19 and 2.54 million in FY19/20.

The registered population at Babylon GP at Hand was 32,393 and 61,273 patients for FY18/19 and FY19/20, respectively. There was total of 60,587 patients after 686 patients registered to the Birmingham site in FY19/20 were removed (Figure 2C). This population included 2563 and 2696 registered patients in FY18/19 and FY19/20, respectively, who received a traditional model of care. This population and the associated spending was removed (Figure 2D).

The remaining 357 practices in the region had an average of 6718 registered patients (range 235-21,688) in FY18/19 and 6943 (range 224-22,969) in FY19/20. There were 48 primary care networks (excluding the Babylon GP at Hand primary care network) in the North West London region, which had on average 49,682 registered patients (range 28,318-80,903) in FY18/19 and 51,358 (range 29,125-83,965) in FY19/20.

Compared with the North West London Collaboration of Clinical Commissioning Group population, the Babylon GP at Hand population was more concentrated in working age adults, had higher rates of employment, and experienced similar levels of deprivation (Table 2).

XSL•FO
RenderX

Table 2. Sociodemographic characteristics of the populations [33,47].

Sociodemographic indicators	FY18/19[a]		FY19/20[b]	
	BGPaH[c] (n=32,394)	NWLCCCG[d] (n=2,398,352)	BGPaH (n=60,587)	NWLCCCG (n=2,478,711)
Population by age band, n (%)				
Female 0-19	729 (2.25)	255,803 (10.67)	987 (1.63)	262,433 (10.59)
Female 20-39	12,221 (37.73)	425,275 (17.73)	24,040 (39.68)	437,143 (17.64)
Female 40-59	1208 (3.73)	294,998 (12.3)	1881 (3.1)	306,834 (12.38)
Female 60-79	205 (0.63)	154,813 (6.45)	244 (0.4)	160,536 (6.48)
Female 80+	37 (0.11)	42,439 (1.77)	39 (0.06)	43,822 (1.77)
Male 0-19	551 (1.7)	269,153 (11.22)	814 (1.34)	275,767 (11.13)
Male 20-39	14,474 (44.68)	430,981 (17.97)	27,757 (45.81)	442,271 (17.84)
Male 40-59	2675 (8.26)	348,327 (14.52)	4436 (7.32)	365,469 (14.74)
Male 60-79	268 (0.83)	146,604 (6.11)	361 (0.6)	153,160 (6.18)
Male 80+	26 (0.08)	29,959 (1.25)	29 (0.05)	31,276 (1.26)
Index of multiple deprivation[e], percentile[f]	N/A[g]	N/A[g]	45th	52nd
Employment status[h] (%)				
Employed[i]	94.2	68.7	90.5	68.9
Unemployed	0.1	5.3	4.0	5.3
Other[j]	5.7	26.1	5.5	25.9

[a]FY18/19: financial years 2018-2019.

[b]FY19/20: financial years 2019-2020.

[c]BGPaH: Babylon GP at Hand practice.

[d]NWLCCCG: North West London Collaboration of Clinical Commissioning Groups.

[e]Available for 97.8% (349/357) of practices.

[f]With the 1st percentile representing the most deprived and the 100th representing the least.

[g]N/A: not available.

[h]Absolute values were not published.

[i]Employed status included “Full-time paid work (30 hours or more each week)”; “Part-time paid work (under 30 hours each week)”; and “Full-time education at school, college or university”.

[j]Other included “Permanently sick or disabled,” “Fully retired from work,” “Looking after the family or home,” and “Doing something else”.

Weighting of Patient Population

The average need index (the factor describing the size of the weighted population relative to the registered population) for Babylon GP at Hand was between 27.5% and 43.9% lower for the core services and the core services adapted weighting methodologies, respectively (Table 3). Smaller differences were observed in the Carr-Hill method.

XSL•FO
RenderX

Table 3. Average need indices for the 3 weighting methodologies for Babylon GP at Hand practice and the average of North West London Collaboration of Clinical Commissioning Groups for FY18/19 and FY19/20.

Weighting methodology	BGPaH[a] need index	NWLCCCG[b] need index	Difference[c], absolute (%)
Core services			
FY18/19[d]	0.66	0.93	0.26 (28.4)
FY19/20[e]	0.67	0.92	0.25 (27.5)
Core services adapted			
FY18/19	0.50	0.88	0.39 (43.7)
FY19/20	0.49	0.88	0.39 (43.9)
Carr-Hill			
FY18/19	1.00	0.93	–0.08 (–8.2)
FY19/20	0.90	0.93	0.03 (3.2)

[a]BGPaH: Babylon GP at Hand practice.

[b]NWLCCCG: North West London Collaboration of Clinical Commissioning Groups.

[c]Numbers may not sum due to rounding.

[d]FY18/19: financial years 2018-2019.

[e]FY19/20: financial years 2019-2020.

Cost Per Weighted Patient

Before statistical analysis, the cost per weighted patient for Babylon GP at Hand was increased by 12% (Figure 2E) to correct for lower than expected activity rates reported in an independent review of the practice [42].

The 1-sided, simple *z* test was performed on a primary care network level, where the spending per weighted patient was normally distributed. A significantly lower cost per weighted patient for Babylon GP at Hand was observed for all weighting methodologies across both years. This was between 12.4% (£93) to 54.4% (£389) lower in FY18/19 and 15.2% (£114) to 50.9% (£362) lower in FY19/20 (Table 4).

Practice-level data were not normally distributed, there was a high number of outliers, and the Babylon GP at Hand practice was not of a comparable size; thus, the 1-sided, simple *z* test was not performed. The Babylon GP at Hand practice's percentile among the other 357 practices is shown Table 5.

Table 4. Summary of cost per weighted patient for Babylon GP at Hand compared with the North West London Collaboration of Clinical Commissioning Groups average, including absolute and percentage differences and 1-sided, simple *z* test results for FY18/19 and FY19/20.

Weighting methodology	BGPaH[a] cost (£) per weighted patient, £	NWLCCCG[b] cost (£) per weighted patient, £	Difference in cost[c] (£), amount (%)	*P* value
Core services				
FY18/19[de]	492	715	–223 (–31.2)	<.001
FY19/20[f]	468	714	–246 (–34.5)	<.001
Core services adapted				
FY18/19[e]	656	748	–93 (–12.4)	.047
FY19/20	635	749	–114 (-15.2)	.006
Carr-Hill				
FY18/19[e]	325	714	–389 (–54.4)	<.001
FY19/20	349	711	–362 (–50.9)	<.001

[a]BGPaH: Babylon GP at Hand.

[b]NWLCCCG: North West London Collaboration of Clinical Commissioning Group average, excluding Babylon GP at Hand.

[c]Numbers may not sum due to rounding.

[d]FY18/19: financial years 2018-2019.

[e]Adjusted for inflation to be comparable to FY19/20 costs.

[f]FY19/20: financial years 2019-2020.

XSL•FO
RenderX

Table 5. Percentile rank of Babylon GP at Hand practice among all practices in North West Central London Collaboration of Clinical Commissioning Groups.

Weighting methodology	Babylon GP at Hand, percentile[a]
Core services	
FY18/19[b]	3rd
FY19/20[c]	3rd
Core services adapted	
FY18/19	15th
FY19/20	9th
Carr-Hill	
FY18/19	2nd
FY19/20	2nd

[a]With the 1st percentile representing the lowest spending per weighted patient and the 100th representing the highest.

[b]FY18/19: financial years 2018-2019.

[c]FY19/20: financial years 2019-2020.

A summary of the total acute spending, registered populations, need indices, weighted populations, and cost per weighted patient for each practice and primary care network can be found in Multimedia Appendix 2.

Discussion

Principal Results

This paper is the first to show that an association between a highly accessible, 24/7, digital-first model of primary care and significantly lower acute hospital costs. This was observed over 2 consecutive years and across all 3 methodologies used to adjust for health care need. The spending per weighted patient for Babylon GP at Hand practice was 12%, 31%, and 54% (£93, P=.047; £223, P<.001; and £389, P<.001) lower than the regional average in FY18/19 for the core services adapted, core services, and Carr-Hill weighting methodologies, respectively. In FY19/20, it was 15%, 35%, and 51% (£114, P=.006; £246, P<.001; and £362, P<.001) lower. This represented a lower total spending for the Babylon GP at Hand population in FY18/19 of £1.37 million, £4.40 million, and £11.6 million for the core services adapted, core services, and Carr-Hill weighting methodologies, respectively. In FY19/20, the equivalent figures were £3.26 million, £9.54 million, and £18.8 million, respectively.

The reduction in hospital care costs observed is likely to be much greater than the additional cost of delivering 24/7, digital-first primary care. In FY19/20, the Babylon GP at Hand practice delivered 23% more appointments per Carr-Hill weighted patient than the national average [35]. Even if primary care costs grew linearly with the number of appointments, this would translate to additional costs of £36 (based on an average funding per patient of £155 in FY19/20 [48] and assuming that primary care funding equals the costs of provision). Even after these additional costs were accounted for, the savings in FY19/20 would still be between £78 and £326 per weighted patient. Furthermore, any additional digital-first primary care costs are borne by the provider (Babylon GP at Hand) and not by the NHS, as NHS primary care practice payments are capitated rather than activity based. The full acute cost savings therefore accrue to the NHS.

Limitations

First, the main limitation is that patient-level data were not available; therefore, it was not possible to examine the causal factors behind the lower costs observed for patients receiving 24/7, digital-first, primary care. Second, given that patients chose which practice to join, there might have been a degree of self-selection that was not corrected for by the weighting formulae used. However, adjustments, such as prior use of health care services by Babylon GP at Hand members, were made for known differences. This was conservative and acted to increase the cost per weighted person for the Babylon GP at Hand practice, suggesting that the cost savings may be greater than those shown. Further work is needed to access patient-level data, which could explain in which areas savings are made, eliminate self-selection bias, and reduce the need for adjustments.

The wider applicability of the findings is limited in part by the registered population of Babylon GP at Hand and by the spending categories returned in the FOI request. The population of Babylon GP at Hand is concentrated in working age adults, 95.9% (58,113/60,587) of patients in FY19/20 were between 20 to 59 years old compared to 62.6% (1,551,717/2,478,709) in the rest of the region. This could partially explain the higher percentage of employment observed in the Babylon GP at Hand population. The weighting formulae adjusted for the population differences, as evidenced by the Core Service Adapted need index in FY19/20 being 43.9% lower for Babylon GP at Hand than the regional average. However, interpretation of the findings is limited for other age ranges. The spending data returned in the FOI request did not include mental health, community, or primary care prescription spending, which represented 36% of the total budget for the Clinical Commissioning Group core services in FY19/20 [36]. Resultantly, the effect of 24/7 digital-first health care on this

XSL•FO
RenderX

portion of spending has not been determined. However, there are reasons to believe that the same effect would be observed in these categories, as has been shown elsewhere [49].

This study is focused on acute hospital spending but did not assess the quality and therefore health care value. Assessing the quality of primary care is difficult given its broadness and the lack of robust quality metrics. However, during the period of investigation, Babylon GP at Hand practice was rated "Good" by the Care Quality Commission; scored 92% and 96% in all available Quality Outcome Framework points in FY18/19 and FY19/20 [50], respectively; and 94.4% (146,077/154,738) of patient ratings for clinical consultations were 4 or 5 stars out of 5 during the study period (SG Winward, MD, unpublished data, April 1, 2018, to March 31, 2020).

The Carr-Hill weighting methodology factors in demographics and other drivers of need, but its purpose is to determine primary care funding rather than acute care spending. This weighting approach was recommended in the FOI response and hence it was included but is not considered as robust as the core services adapted and core services methodologies. Therefore, the central finding of this paper is a 15%-35% lower spending per weighted patient for members of Babylon GP at Hand in FY19/20.

The accuracy of the analysis in this paper is contingent on the quality and reliability of the NHS data that were provided in the FOI request (Multimedia Appendix 1).

Comparison With Prior Work

To our knowledge, this study is the first to assess the impact that a highly accessible, digital-first model of primary care has on acute hospital spending. Although the findings cannot infer causality, they are consistent with those of other publications. These include, for example, a link between accessible primary care and reduced demand for other services [32]; and an NHS England commissioned report on Babylon GP at Hand, which found patients were significantly less likely to attend ED and outpatient appointments than was a control population [42]. The authors of the report have since called for more evidence to be obtained on the sustainability of such services, as this paper seeks to provide [51].

The paper is aligned with the majority of the literature in showing that digital health solutions can reduce costs [23-25,27]. It adds insight in two areas where research is sparse. First, it is an assessment of a digital-first model of care for a wide range of conditions as opposed to a digital tool in a single condition; second, it quantifies cost savings from the perspective of the health system [19]. The results are particularly important in the context of the national direction of travel laid out in the NHS Long Term Plan, which states that all patients must be able to access digital-first primary care by the end of the 2024 financial year [31].

Several areas requiring further research have been identified. To increase confidence in the conclusions, assessment of patient-level data over all spending categories (ie, mental health, prescriptions, and community and acute hospital spending) during the same period is required. Further research is also needed to fully assess the impact of the model of care on quality outcomes. The effectiveness of telehealth solutions has been shown to be linked to the provider [27], so further work with alternate providers is required to establish if the observed benefits are uniform.

Conclusions

This paper has demonstrated that highly accessible, 24/7, digital-first primary care was associated with lower acute hospital spending for a health system. This effect was sustained over a 2-year period, during which the population under investigation doubled in size, demonstrating that the effect is scalable. Further work using patient-level data is needed to be able to generalize these findings to a wider demographic of patients and to understand the efficacy of digital-first primary care across different populations of patients.

Authors' Contributions

SW contributed to the conceptualization, methodology, formal analysis, and writing of the original draft. TP contributed to the writing and review of the manuscript. MAS contributed to the writing, review, and editing of the manuscript. MN provided supervision and contributed to the writing and review of the manuscript.

Conflicts of Interest

All the authors are employees of, and have share options in, Babylon Health.

Multimedia Appendix 1
Freedom of Information (FOI) request and response.
[XLSX File (Microsoft Excel File), 20 KB - jmir_v23i7e24917_app1.xlsx]

Multimedia Appendix 2
Summary of results.
[XLSX File (Microsoft Excel File), 155 KB - jmir_v23i7e24917_app2.xlsx]

References

1. Current health expenditure as percentage of gross domestic product - data by WHO region. World Health Organization. URL: https://apps.who.int/gho/data/view.main.GHEDCHEGDPSHA2011REGv?lang=en [accessed 2020-11-13]

XSL•FO
RenderX

2. United Nations: Department of Economic and Social Affairs. World Population Ageing 2019. 2020. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_2019_worldpopulationageing_report.pdf [accessed 2020-11-13]
3. Nghiem SH, Connelly LB. Convergence and determinants of health expenditures in OECD countries. Health Econ Rev 2017 Aug 17;7(1):29 [FREE Full text] [doi: 10.1186/s13561-017-0164-4] [Medline: 28819772]
4. Ricciardi W, Pita Barros P, Bourek A, Brouwer W, Kelsey T, Lehtonen L, Expert Panel on Effective Ways of Investing in Health (EXPH). How to govern the digital transformation of health services. Eur J Public Health 2019 Oct 01;29(Supplement_3):7-12 [FREE Full text] [doi: 10.1093/eurpub/ckz165] [Medline: 31738442]
5. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva: World Health Organization; 2016.
6. Bashshur RL, Shannon G, Krupinski EA, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013 May;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]
7. Expert Panel on effective ways of investing in Health. Assessing the impact of digital transformation of health services. European Commission. 2019 Jan. URL: https://ec.europa.eu/health/sites/health/files/expert_panel/docs/022_digitaltransformation_en.pdf [accessed 2020-11-13]
8. Baird B. How has general practice responded to the Covid-19 (coronavirus) outbreak?.: Kings Fund URL: https://www.kingsfund.org.uk/blog/2020/04/covid-19-general-practice [accessed 2020-11-13]
9. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
10. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Apr 02:382. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
11. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
12. Telehealth. World Health Organization. URL: http://www.who.int/gho/goe/telehealth/en/ [accessed 2020-11-13]
13. Powell RE, Stone D, Hollander JE. Patient and health system experience with implementation of an enterprise-wide telehealth scheduled video visit program: mixed-methods study. JMIR Med Inform 2018 Feb 13;6(1):e10 [FREE Full text] [doi: 10.2196/medinform.8479] [Medline: 29439947]
14. Fatehi F, Martin-Khan M, Smith AC, Russell AW, Gray LC. Patient satisfaction with video teleconsultation in a virtual diabetes outreach clinic. Diabetes Technol Ther 2015 Jan;17(1):43-48. [doi: 10.1089/dia.2014.0159] [Medline: 25296189]
15. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 Dec;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
16. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Dec 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
17. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
18. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
19. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298. [doi: 10.2196/17298]
20. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010 Jun 16;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
21. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]
22. McLean S, Nurmatov U, Liu JL, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011 Jul 06(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
23. Elbert NJ, van OH, van RW, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
24. Polisena J, Coyle D, Coyle K, McGill S. Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations. Int J Technol Assess Health Care 2009 Jul;25(3):339-349. [doi: 10.1017/S0266462309990201] [Medline: 19619353]
25. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]

XSL•FO
RenderX

26. Flodgren G, Rachas A, Farmer A, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
27. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
28. Rahimi K. Digital health and the elusive quest for cost savings. Lancet Digit Health 2019 Jul;1(3):e108-e109 [FREE Full text] [doi: 10.1016/S2589-7500(19)30056-1] [Medline: 33323258]
29. Sweet CC, Jasik CB, Diebold A, DuPuis A, Jendretzke B. Cost savings and reduced health care utilization associated with participation in a digital diabetes prevention program in an adult workforce population. J Health Econ Outcomes Res 2020;7(2):139-147 [FREE Full text] [doi: 10.36469/jheor.2020.14529] [Medline: 32884964]
30. Whaley CM, Bollyky JB, Lu W, Painter S, Schneider J, Zhao Z, et al. Reduced medical spending associated with increased use of a remote diabetes management program and lower mean blood glucose values. J Med Econ 2019 Sep;22(9):869-877. [doi: 10.1080/13696998.2019.1609483] [Medline: 31012392]
31. NHS Long Term Plan. The National Health Service. 2019 Jan 07. URL: https://www.longtermplan.nhs.uk [accessed 2020-11-13]
32. Whittaker W, Anselmi L, Kristensen S, Lau Y, Bailey S, Bower P, et al. Associations between extending access to primary care and emergency department visits: a difference-in-differences analysis. PLoS Med 2016 Sep;13(9):e1002113 [FREE Full text] [doi: 10.1371/journal.pmed.1002113] [Medline: 27598248]
33. Patients registered at a GP practice. NHS Digital. URL: https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice [accessed 2020-11-13]
34. Standard General Medical Services Contract. NHS England and NHS Improvement. 2019. URL: https://www.england.nhs.uk/wp-content/uploads/2019/11/gms-contract-october-2019.pdf [accessed 2020-11-13]
35. Appointments in general practice (2018-2020). NHS Digtial. URL: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice [accessed 2020-11-13]
36. Infographics: fair shares - a guide to NHS allocations. NHS England and NHS Improvement. 2020 Feb 10. URL: https://www.england.nhs.uk/publication/infographics-fair-shares-a-guide-to-nhs-allocations/ [accessed 2020-11-13]
37. Technical guide to allocation formulae and pace of change for 2019/20 to 2023/24 revenue allocations. NHS England. 2019 May 30. URL: https://www.england.nhs.uk/publication/technical-guide-to-allocation-formulae-and-pace-of-change-for-2019-20-to-2023-24-revenue-allocations/ [accessed 2020-11-13]
38. Dyson M. Global sum allocation formula. British Medical Association. URL: https://www.bma.org.uk/advice-and-support/gp-practices/funding-and-contracts/global-sum-allocation-formula [accessed 2020-11-13]
39. NHS payments to general practice. NHS Digital. URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-payments-to-general-practice [accessed 2020-11-13]
40. Baird B. Primary care networks explained. Kings Fund. 2019. URL: https://www.kingsfund.org.uk/publications/primary-care-networks-explained [accessed 2020-11-13]
41. Primary care networks. NHS England. URL: https://www.england.nhs.uk/primary-care/primary-care-networks/ [accessed 2020-11-13]
42. Evaluation of Babylon GP at Hand Final report. Ipsos MORI and York Health Economics Consortium. 2019 May. URL: https://www.hammersmithfulhamccg.nhs.uk/newspublications/publications-policies.aspx?n=9446 [accessed 2020-11-13]
43. Hospital admitted patient care activity. NHS Digital. URL: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity [accessed 2020-11-13]
44. National cost collection for the NHS. NHS England. 2019. URL: https://www.england.nhs.uk/national-cost-collection/ [accessed 2020-11-13]
45. GDP deflators at market prices, and money GDP November 2020 (Spending Review). Office of National Statistics, GOVUK. URL: https://www.gov.uk/government/statistics/gdp-deflators-at-market-prices-and-money-gdp-november-2020-spending-review [accessed 2020-11-13]
46. Jones E, Oliphant T, Peterson P. SciPy: open source scientific tools for Python. 2001. URL: http://www.scipy.org [accessed 2020-11-13]
47. Public health profiles. Public Health England. 2021. URL: https://fingertips.phe.org.uk [accessed 2020-11-13]
48. GP practices were paid an average of £155 per patient in 2019/20. Pulse Today. 2020. URL: https://www.pulsetoday.co.uk/news/practice-personal-finance/gp-practices-were-paid-an-average-of-155-per-patient-in-2019-20/ [accessed 2020-11-13]
49. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The empirical evidence for telemedicine interventions in mental disorders. Telemed J E Health 2016 Feb;22(2):87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]
50. Quality and outcomes framework. NHS Digital. URL: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data [accessed 2020-11-13]
51. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]

XSL•FO
RenderX

Abbreviations

ED: emergency department
FOI: freedom of information
FY: financial year
GDP: gross domestic product
NHS: National Health Service (of the United Kingdom)

Edited by R Kukafka; submitted 12.10.20; peer-reviewed by YH Kwan, X He; comments to author 03.11.20; revised version received 27.11.20; accepted 04.06.21; published 22.07.21.

Please cite as:
Winward S, Patel T, Al-Saffar M, Noble M
The Effect of 24/7, Digital-First, NHS Primary Care on Acute Hospital Spending: Retrospective Observational Analysis
J Med Internet Res 2021;23(7):e24917
URL: https://www.jmir.org/2021/7/e24917
doi:10.2196/24917
PMID:34292160

©Sam Winward, Tejal Patel, Mazin Al-Saffar, Matthew Noble. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 22.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Sharing Economy in China's Aging Industry: Applications, Challenges, and Recommendations

Yaolin Hu[1], MA; Jian Wang[1,2], Prof Dr; Stephen Nicholas[3,4,5,6], Prof Dr; Elizabeth Maitland[7], Prof Dr

[1]Dong Fureng Institute of Economic and Social Development, Wuhan University, Wuhan, China

[2]Center for Health Economics and Management, Economics and Management School, Wuhan University, Wuhan, China

[3]Australian National Institute of Management and Commerce, Sydney, Australia

[4]Newcastle Business School, University of Newcastle, Newcastle, Australia

[5]School of Economics and School of Management, Tianjin Normal University, Tianjin, China

[6]Research Institute of International Strategies, Guangdong University of Foreign Studies, Guangzhou, China

[7]School of Management, University of Liverpool, Liverpool, United Kingdom

Corresponding Author:
Jian Wang, Prof Dr
Dong Fureng Institute of Economic and Social Development
Wuhan University
299 Bayi Road, Wuchang District
Wuhan, 430072
China
Phone: 86 13864157135
Fax: 86 010 66212101
Email: wangjian993@whu.edu.cn

Abstract

Background: All aging societies face the challenge of allocating limited resources for the highest value of use. The sharing economy provides one method to address the imbalance between the demand and supply of health services to the older adult population. With a substantial aging population, China's practices in the sharing aging industry may set examples for other "getting old before getting rich" countries.

Objective: There is a gap in both the data and research on China's aging industry sharing economy. This paper addresses these data and research lacunae by constructing a framework for the application of a sharing model in China's aging industry, by assessing the current state of the aging industry sharing economy, by setting out the challenges to the sharing aging health care and service economy, and by making recommendations for the development of the aging industry sharing economy.

Methods: This paper constructs a sharing economy framework in the aging industry covering four aspects (*people*, *facilities*, *capital*, and *information*) to test the current state and future prospects of China's aging industry sharing economy.

Results: In people sharing, we analyzed the sharing of emotional companionship, doctors, nurses, nursing attendants, and domestic helpers. We discussed facility sharing models from the point of land and housing, medical devices, and other items such as pensioner meals and shared medicine bins. We acknowledge that crowdfunding platforms have developed fast in China, but many older adult users faced problems in their operation. Information sharing is a developing field, which can optimize users' experiences and should help older adults filter out misinformation, but China currently does not have adequate sharing information platforms for older adults.

Conclusions: We identified four major challenges in China's aging industry sharing economy: poor adaptability to technology for older adults, mediocre quality of shared services, *one-size-fits-all* and the concept of the *useless elderly*, and shortage of qualified practitioners. We make recommendations for specific measures by governments, communities, and enterprises to improve the sharing economy in the aging industry.

(J Med Internet Res 2021;23(7):e27758) doi:10.2196/27758

KEYWORDS

sharing economy; aging industry; older adult population; older adults; aging; health services; sharing model; sharing; China; East Asia

XSL•FO
RenderX

Introduction

Population Aging

Population aging is a global phenomenon, with many aging societies such as western Europe, Japan, and the United States amassing experience in aging social governance. China is a typical resource-limited country that is "getting old before getting rich," with 176 million or 12.6% of the population older than 65 years [1]. According to the United Nation's criteria, China has been an aging country since 2000, with the older-aged population likely to exceed 300 million by 2025 as China transitions from a mildly aging country to a moderately aging one [2]. Only in 1994 did the Chinese government address the escalating population aging problem with a *Seven-year Development Program for China's Work on Aging (1994-2000)*. The magnitude of the aging challenge is reflected in China's gross domestic product (GDP) rank as the second in the world but its per capita GDP, estimated at US $10,261 in 2019, ranking only 80th in the world [3]. It is estimated that 600 million people in China earn only ¥1000 (about US $152) per month, with high poverty levels concentrated in the old aged [4]. China's aging population has brought about complex social problems, including the demand for age care facilities and services, putting the government under severe governance and financial pressure.

The Sharing Economy in the Aging Industry

Sharing Economy

Many problems brought about by population aging, such as shortage of health care staff, uneven distribution of aged care facilities, and poor services to the aged population, reflect the failure of age care resource supply to meet age care demand. Given the fixed and inflexible nature of many age care resources, the sharing economy's innovations in technology and management promise to improve the balance between age care demand and supply. Faced by an absence of a consensus on the definition of the sharing economy and little published quantitative or qualitative research on the sharing aging industry [5,6], we specify an aging industry sharing economy framework, which is assessed and evaluated for China's sharing aging industry. Organizations in the sharing economy have the following characteristics: they are organized as digital platforms enabling offline transactions between users, they facilitate peer-to-peer transactions and business-to-business operations, they emphasize temporary access rather than ownership, and they focus on the use of underused or undersupplied resources [7]. The sharing economy, as part of the digital economy [8], has transformed services, transport, finance, and education, with the health care industry starting to apply sharing models [9]. Sharing models match information on the supply and demand for resources with the help of internet technology. One of the most important features and influences of the sharing economy has been how it has changed people's views on resources from ownership to the temporary right to access [10], which promotes the circulation of idle, misallocated, or underused resources and improves the efficiency of resource use. In 2019, the transaction volume of China's sharing economy was estimated at ¥3.3 billion (US $509.5 million), comprising over 800 million citizens [11].

Aging Industry and Challenges to the Health Care Sharing Economy

The aging industry is a general name for all relevant industrial sectors that provide products and services to citizens in their old age and younger people preparing to enter old age, such as the wealth management and superannuation industries. China's traditional culture of family-based old age care is being re-evaluated, as the middle-aged population consider their future old age care and their care of older-aged family members. Despite the aggressive innovation in digital platforms to manage the supply of products and services, provision of old age services remains unresolved in China. First, the development of the aging industry has not paid sufficient attention to the economic differences within the aging population in China. There has been a bifurcation between old age consumers of high-end health care products and services, not available to all, and low-end health care products consumed by less economically well-off older-aged consumers. Second, a *one-size-fits-all* approach has been adopted by most age health providers, with insufficient attention paid to the diverse physical and mental states of the older-aged consumer population, ignoring the differences between the younger elderly and older elderly. Third, there has been insufficient focus on resource supply shortages in the aging industry, especially nursing attendants, nursing homes, and other infrastructure.

The sharing economy promises the ability to maximize the value of underused and misallocated resources, and to address supply shortages, providing solutions to rebalance the supply and demand of aging industry resources. The Chinese government has started to address the underuse and undersupply of resources in the aging industry [12,13], supporting and guiding the development of the sharing economy with the aim to improve the efficiency of social resource use. In 2016, the *13th Five-year Plan for the Development of Civil Affairs* encouraged mutual assistance and coordinative efforts in the provision of old age service delivery. Although some enterprises in China's aging industry have adopted the sharing and mutual assistance model, the sharing model in the medical treatment industry concentrates on serving people of all ages. Surprisingly, given the large size of the aged population, data and research on China's aging industry sharing economy have been insufficient. We address these lacunae by constructing a framework for the application of the sharing model in China's aging industry, by assessing the current state of the aging industry sharing economy, by setting out the challenges to the sharing aging health care economy, and by making recommendations for the development of the aging industry sharing economy.

Methods

There has been little research on the sharing economy in the aging industry in China [6]. Depending on actual qualitative data from China's sharing aging industry, our grounded method aims to explain the current status (the status quo) of an institution, China's sharing aging industry, and its implications (challenges and applications). Following the grounded qualitative method, we collected and assembled data and information from a wide range of aging industry government,

XSL•FO
RenderX

commercial and public sources and websites, academic papers, public reports, institutional reports, and fieldwork. The first analytical step was to categorize in Textbox 1 the types of services and activities applicable to the sharing economy model for older people. Second, we used the qualitative data in Textbox 1 to specify in Figure 1 an analytical framework for assessing the aging industry in China. Finally, we summarized, interpreted, and critically evaluated the information in Textbox 1 to explicate the sharing aging industry status quo and set out the challenges and implications, providing a range of recommendations.

In a 2016 study of sharing services for those 85 years or older, Ward and Coughlin identified meals, medication management, transportation, housekeeping, recreation and wellness activities, security, and personal care as elements in the sharing economy [14]. In China, the National Committee on Aging researched urban home-based older people's satisfaction with domestic services, nursing services, chatting services, and legal aid services [15], finding these four aspects of old age services under substantial pressure. The Beijing Municipal Civil Affairs Bureau found that urban older people had little space for outdoor activities, but the opposite was true for rural older people [16]. Besides accepting services in their old age, older people were also willing to contribute to society in their old age. It was found that 45.6% of older adults (over 100 million people) often participate in various public welfare activities [17]. Based on government and public reports, websites, and academic papers, Textbox 1 categorizes the types of services and activities applicable to the sharing economy model for older people, both in terms of the status quo and prospective improvements for the future.

To conceptualize and operationalize the service categories in Textbox 1, Figure 1 provides a comprehensive sharing economy framework for the aging industry in China, comprising people, facilities, capital [9], and information.

In the next section, we assess the key elements in the aging industry sharing economy within a *people*, *facilities*, *capital*, and *information* framework in Figure 1.

XSL•FO
RenderX

Textbox 1. Services that could use the sharing economy model needed by older people from government, commercial and public welfare sources and websites, academic papers, public reports, institutional reports, and fieldwork.

Basic living needs

- Houses
  - **Status quo**
  - High urban housing prices make it difficult for children to house parents.
  - Inadequate care for older parents due to children's work requirements
  - Most housing design does not cater for older adults' needs
  - The rural older adults live alone with large housing space because of young people's immigration to cities, but the health care facilities are poor
  - Some cities developing housing sharing services for older adults
  - **Prospect**
  - Need for better home care
  - Need for specialist housing for older adults
  - The rural areas share housing resources with the urban areas
- Outdoor activity space
  - **Status quo**
  - Scarce urban outdoor activity space for older adults
  - Abundant rural outdoor space, but the activity facilities are poor
  - **Prospect**
  - Communities can develop special activity space for older residents.
  - By sharing idle land resources in rural areas, older adults in both urban and rural areas can get high quality activity space.
- Housekeeping
  - **Status quo**
  - Sharing platforms for peer-to-peer housekeeping services need improvement.
  - **Prospect**
  - Better sharing in housekeeping services can improve the life quality of home-based older people.
- Meals
  - **Status quo**
  - Few specialist takeaway services catering to the health needs of older adults
  - There have been sharing chefs to cook in users' homes.
  - **Prospect**
  - Need to expand sharing economy's provision of nutritious meals suitable for older adults
- Transportation
  - **Status quo**
  - Sharing transportation platforms in China (eg, Didi) do not cater for the travel needs of older adults.
  - **Prospect**
  - Sharing transportation platforms for older adult

Medical and nursing needs

- Nursing care
  - **Status quo**
  - Shortage of professional nursing services for older adults at home

XSL•FO
RenderX

- **Prospect**
- Sharing model providing efficient allocation of nursing resources for older adults at home

- Medication management
  - **Status quo**
  - Difficult for older adults to access limited high quality medical resources
  - Older adults need web-based systems for reminder and renewal of medicine prescriptions.
  - **Prospect**
  - Internet health care on access and adaptation

- Personal care
  - **Status quo**
  - Good information for older adults on medical data
  - **Prospect**
  - Internet medical care providing enhanced access to professional health service for older adults

Emotional and spiritual needs

- Recreation and chatting and company
  - **Status quo**
  - Poor level of entertainment and spiritual needs of older adults
  - **Prospect**
  - A sharing economy provides improved entertainment and spiritual companionship for older adults.

- Work
  - **Status quo**
  - Many older people still want to participate in work.
  - **Prospect**
  - Develop more sharing platforms to help older adults get job information.

Results

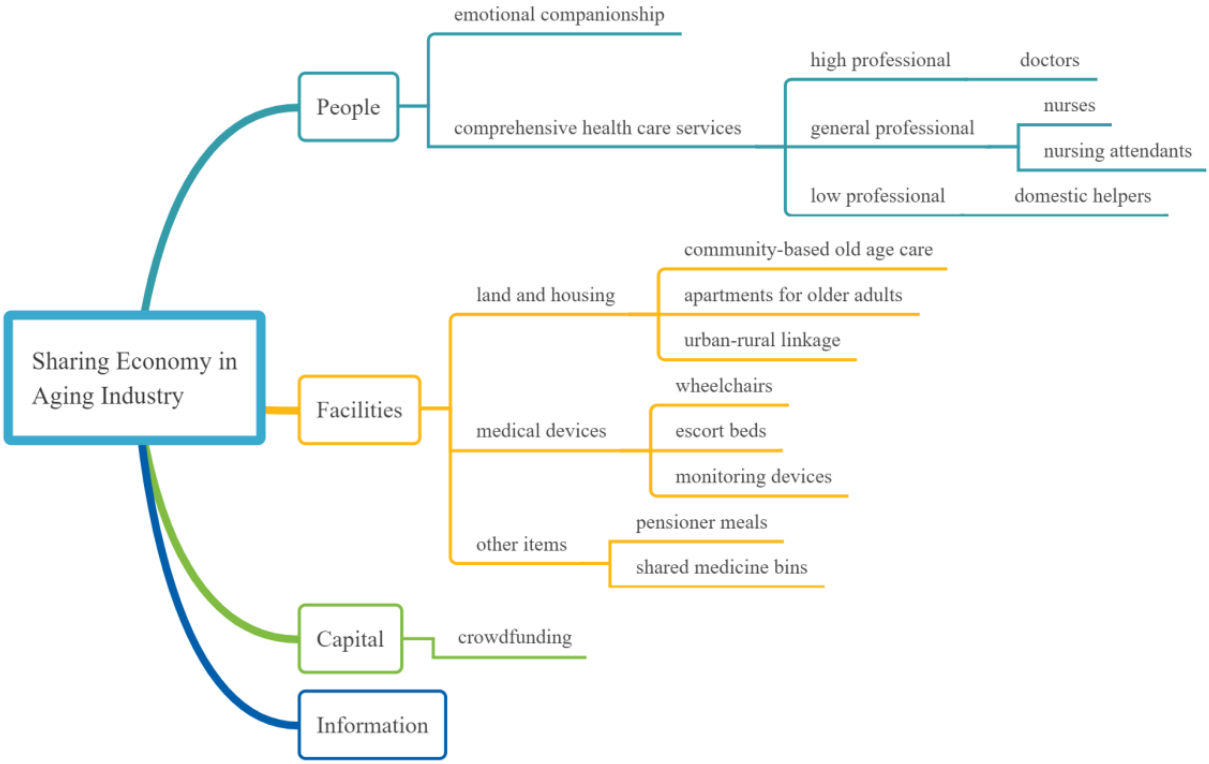
People

In the sharing economy model in Figure 1, sharing *people* refers to the supply side provision of people's time, expertise, skills, or emotional company to old age consumers. Many old people taking care of themselves at home face the challenges of self-care ability, loneliness, and a feeling of being neglected [18]. Family planning, including the one-child policy and economic pressures on family size, have constrained the size of the average Chinese family to 3.1 persons in 2010, limiting family care for older family members [1]. In addition, the heavy influence of the Confucian tradition of *filial piety* is eroding among young people, especially those who have migrated from rural to urban areas, have less financial capacity to house, and have less time to meet the physical and emotional needs of old age family members. As shown in Figure 1, sharing of people can be divided into pure emotional companionship and comprehensive health care services, comprising high professional level, general professional level, and low professional level.

China has innovated in sharing people for emotional companionship, such as the *sharing children* model in Lianhu District, Xi'an City and the *caring tenants* project in Wuhan City. Supported by the community and funded by public welfare funds, Lianhu's 2019 *good-neighbor sharing children* project recruited college students, with the consent of the old age person's family, to share older persons' extra rooms at a low price or free of charge. College students provided companionship, life care, and other services. Wuhan's *caring tenants* project is similar, providing young people free-of-charge house sharing in return for daily companionship, with an adaptation period set before signing the mutual assistance agreement. In both projects, college students or young adults living with the old aged did not provide professional services but provided social engagement [19] and served the older aged emotional needs through companionship in their daily life.

XSL•FO
RenderX

Figure 1. The framework of the sharing economy in the aging industry.



Comprehensive health services relied on the expertise and skills provided by professional health specialists, divided into high professional, general professional, and low professional levels (Figure 1). The shared service suppliers at high professional level were mainly doctors, whose shared services in China has mainly been through telemedicine. Shared telemedicine services address the shortages of graded diagnosis and treatment, which are mainly provided at tertiary high-level hospitals. Older patients, especially those without accompanying family members, usually have poor access to timely tertiary hospital medical services, both due to their poor health constraining visits to a tertiary hospital and because their age and education level constrain their understanding of treatment choices between different hospital levels. Older patients with multimorbidities undertake many hospital visits, where many doctors prescribe only standardized, rather than personalized, chronic disease medicines [20]. Finally, hospital visits pose high risks of cross infections to older patients, especially those with low immunity to diseases. Telemedicine addresses these issues.

Telemedicine has been growing rapidly since 2000 in China, with internet medical platforms such as 39Health.com and DXY. Initially, user activity was not high, and the platforms were subject to restrictive policy controls. Currently, there are more than 900 internet hospitals in China, coupled with a telemedicine collaboration network covering more than 24,000 medical institutions in prefecture-level cities and more than 5500 hospitals at the secondary level or above that are capable of online services [21]. The COVID-19 outbreak has made hospitals more aware that telemedicine can help avoid cross infection, promote graded diagnosis and treatment, and make medical resources more accessible. During COVID-19, government authorities issued a number of official directions to accelerate the development of internet medical treatment. For example, the Wuhan Healthcare Security Administration launched medical insurance payment for the WeDoctor Internet General Hospital in early 2020, pioneering the formation of an online payment loop for diagnosis and treatment. By June 2020, the number of online medical users had reached 276 million, accounting for 29.4% of China's total internet users [22]. In addition, during COVID-19, internet users purchased 26.4% of their medicine, health apparatus, and other medical supplies online, and 17.9% of the users accessed online medical services such as online registration and consultation [22]. Internet diagnosis and treatment apps allowed users to choose doctors according to their diseases, and doctors used texts, audio, and videos to diagnose online patients and prescribe treatments and medicine.

One example of an internet diagnosis and treatment app, Good Doctor covers 9852 public hospitals in China [23], with many well-known doctors accessed online when face-to-face appointments were hard to make. Importantly, apps such as Good Doctor do not require older-aged patients to have particularly high mobile phone operation skills, with these platforms providing voice input, voice recognition, and online phone calls for old age patients who cannot type texts on their mobile phones. Since the introduction of the *family doctor* policy in China, most of the online health platforms provide the option of signing a long-term contract for online consultation, prescription, and health consultation.

Doctor sharing models also include the combination of medical treatment and old age care, which benefit residents of old age care facilities [24]. In China, every thousand people had only 2.8 doctors (including assistant doctors) and 3.18 nurses in 2019

XSL•FO
RenderX

[25]. There were only 31.6 nursing beds per 1000 older-aged persons, but the nursing bed vacancy rate in some institutions was 50% to 60% [26]. Doctor sharing allows underused resources in old age care facilities to accept semidisabled or completely disabled older-aged people who are seriously ill. For example, to avoid older patients crowding out medical resources by using hospitals as nursing homes, doctor sharing models promote the combination of medical treatment and old age care within nursing homes and improve referral and cooperation between hospitals and nursing homes, maximizing both care and bed capacity. To promote doctor sharing, China's National Health Commission and the World Health Organization launched 199 best practice examples of combining medical treatment and age care in 2019 [27]. One example is Beijing Longfu Hospital that combines apartments for older adults as a branch of the hospital, housing consulting rooms, and long-term doctor and nurse services to more than 200 old age apartment residents. In addition, the Longfu Hospital cooperated with external old age care institutions through online appointments and treatments.

In Figure 1, service providers at the general professional level include nurses and nursing attendants. In the doctor sharing model previously discussed, nurses and doctors work together to provide medical services to the old aged. The most typical model of nurse sharing in China is to reserve a nurse's "door-to-door service" through an app. Existing nurse sharing apps such as Homeincare, U-nursing, and Goldnurse provide nurse services such as catheterization, bedsore nursing, stoma care, enema care, and sputum aspiration as well as companionship during consultations. The design concept for nursing sharing apps is similar to Uber, allowing nurses to take appointments outside the regular 9 to 5 working hours and allowing users to find suitable nurses who can provide door-to-door services based on their previous service evaluation results.

Although nursing assistants also provide everyday life care services for the old aged, there are constraints on the emergence and use of nurse attendants sharing models. First, there is a severe shortage of nursing attendants. In Beijing, for example, there are about 3.92 million people 60 years or older with Beijing household registration but only 15,000 nursing attendants [28]. Second, nursing attendants vary in competency. An access system for nursing attendants to serve old age patients was cancelled in 2017, with the government requiring the nursing institutions to strengthen nurse attendant training. Third, the social recognition of nursing attendants is not high, with many urban families hiring a nanny to undertake the nursing attendant role, and many rural families regard nursing attendants as an unnecessary luxury. Internet platforms have not addressed the imperfect nursing attendant market in China, which leaves many vulnerable older people without low-level daily care.

Service providers at the lowest professional level in Figure 1 are mainly domestic helpers, who undertake housework, cooking, and shopping for the old aged. Driven by internet technology, China's domestic service market has formed a relatively mature online-to-offline (O2O) model, allowing users to make appointments and payments, leave comments, and rate domestic helpers through apps. One of China's earliest O2O platforms, 58.com app, provides point-to-point housekeeping service information, optimizing the allocation of the domestic service labor force. China has also experimented with *time bank* platforms, where service providers do not get paid for their services but deposit in a time bank their service time, which they can use to pay for services they consume. One example is Crown Community of Yanliang District that organized appointments with workers willing to provide housekeeping services for the older adults through mobile phones. The service providers recorded their service time in the community's time bank, which they could *withdraw* when they, or their older relatives, required help. In 2018, the Ministry of Civil Affairs of China explicitly included *time bank* in the nationwide pilot reform of home-based community old age care services, hoping to establish an operational model that could be promoted throughout the country [29]. So far, the time bank model has not been widespread in China.

Facilities

Sharing of facilities is the primary objective in the sharing economy, with Airbnb and Uber the best-known examples. In China, sharing facilities for the old aged include land and housing, medical devices, and other items. In Figure 1, China's aging industry can be divided into community-based old age care, apartments for older adults, and urban-rural linkage from the point of land and housing.

Community-based old age care is an extension and supplement to home-based care. Since the 1980s, communities have become the most important basic unit of social governance in China, which played a central role in the prevention and control of COVID-19 in 2020. Community-based old age care services are available in most cities, typically involving a shared space at day care centers. Although not sharing land and buildings, virtual nursing homes have been established in communities boasting good communications infrastructure, where the virtual community relies on a unified information management platform to meet older adult care needs online. For example, the Home-based Elderly Care Center in Suzhou is a virtual communications platform providing nursing home information through a *Fun-at-Home* older adult care service system, comprising 53 services in 6 categories, including housekeeping, property maintenance, and health care, such as companionship in hospital, entertainment and learning, and emergency help. Services are offered by professional management companies, and the platform allows older age users to order and evaluate services over the phone. The community plays an important role as a resource coordinator in both the establishment of day care centers and the virtual nursing home based on information technology. By integrating the community's public space and older adult care service resources, community-based older adult care enables older adults to enjoy the benefits of home-based and professional older adult care at the same time.

Apartments for older adults are not nursing homes. Based on the experience of countries with apartments for the old aged in the United States, Japan, and Canada, apartments for the old aged provide an organizing body mainly for social functions, rather than social governance, and operate according to market rules. The construction of older adult apartments in China is

XSL•FO
RenderX

currently showing uneven development. Although high-end apartments for older adults are in sufficient supply, due to high prices and low demand, low-end apartments, mainly funded by government financial aid and serving disadvantaged people, are in short supply. China is currently promoting the construction of low-end apartments for older adults and applying the concept of the sharing economy in the process. First, in terms of construction funds, the government mainly cooperates with private capital through the private-public partnership (PPP) model, rather than relying solely on government financial funds. Second, apartments for older adults are generally built in areas where living space is scarce or older adult care resources are insufficient. Third, jointly built by the government and private capital, low-end apartments for older adults often prompt hospitals in the vicinity to integrate their medical services with apartment-based older adult care. Government's participation means that the price is relatively low, with Beijing's first PPP apartment for older adults charging a housing fee plus basic service fee of only ¥4000 to ¥8000 (US $618-$1237) per month versus the commercial rate of ¥6000 to ¥16,000 (US $927-$2473) [30]. Through the integration of older adult care resources, this type of apartment improves the quality of older adult care services while also alleviating the financial pressure on older adults.

Urban-rural linkages in Figure 1 impact the old age sharing economy through rural-urban migration, a large rural aged population, poor rural old age care infrastructure including hospitals, and a persistent urban-rural income gap. Older adults in cities enjoy access to better medical resources and older adult care services than rural older adults, but urban areas boast less land and housing space for old age care facilities. According to a survey by the Beijing Municipal Civil Affairs Bureau on the home-based older adult care facilities in the urban area of Beijing, 11.8% of the registered communities had no outdoor activity space, 28.5% had only one outdoor activity site, and 23.6% had two or more outdoor activity sites [16]. Differences within cities and between urban-rural communities in land and residential space shapes the sharing economy. For example, more than 2000 farms in the Beijing municipality have joined a rural-urban sharing program, allowing farmers to post information about their idle land and houses on an app that other citizens can rent. Through farm sharing, not only can farmers increase their income, but older adults who have been living in cities for a long time and yearn for a rural life can expand their living experience. In Nanjing, to address the shortage of rural older adult care resources, professional nursing institutions have transformed idle homes of farmers into apartments for older adults and provided medical and nursing services at the same time.

Under facilities in Figure 1, medical devices are part of the sharing economy, although not all medical devices can, or should, be shared. Medical devices such as wheelchairs and escort beds can be accessed through mobile phones in a manner similar to the bike sharing model, and monitoring equipment is shared under the direction of medical staff. However, the sustainability of sharing medical equipment is likely limited, since the average price of a manual wheelchair is about ¥300 (US $46), making them affordable for families with immobile patients who require long-term wheelchair use. The sustainability of sharing escort beds is also questionable. Foldable beds for accompanying family members are an economical and efficient alternative to an escort bed. In addition, many hospitals provide accommodation services for accompanying family members. The sharing of monitoring devices is more like a self-service, where hospitals provide weight scales and blood pressure monitors at the hospital entrance for free use.

Additionally, under facilities in Figure 1, other items for older adult's daily use are different from these items used by other age groups, such as food, which impacts China's high rate of cardiovascular disease and cancer [31]. As previously discussed, the people sharing model allows older age people to make appointments for domestic helpers to cook home meals. Although China's takeaway market focuses on younger people's demand for fast food, there is scope for expanded *pensioner meals*, already provided by some restaurants, and for specialist takeaway meal providers to exclusively provide food for older adults. The Guangzhou government launched a program to deliver nutritious meals to older adults in 2016 [32], but the takeaway food market for the older aged requires further development. There are also examples of drug sharing models in China. *Shared medicine bins*, which work much like the vending machines, can be found in some communities, which dispense common medicines such as cold medicine, analgesic, and anti-inflammatory drugs. Users can pay at the shared medicine bin using their mobile phones.

Capital

With more than 95% of the population covered by medical insurance since 2013, general health coverage is not a major issue requiring help from the sharing economy [33]. However, the small percentage of the noninsured population can use crowdfunding platforms, such as Shuidichou, Qingsongchou, and Aixinchou. For example, Shuidichou, founded in 2016, has raised more than ¥20 billion (US $3.09 billion) for patients with serious diseases from more than 250 million donors. These crowdfunding platforms are currently facing development difficulties in China. Although these crowdfunding platforms can fund people in financial need, they also face issues in their operation, such as false information, unregulated funds, and advertising harassment, with most crowdfunding platforms operated by profit-oriented companies.

Information Sharing

In Figure 1, information sharing, also called aggregation models in China, is the final element in our sharing aging industry model. Aggregation models refer to the simultaneous access to many other platforms via one platform, where information sharing is not limited to a single platform. In 2017, the navigation software AMAP took the lead in connecting links of different bicycle and car sharing providers on its navigation page, and WeChat now provides free access to takeaways, sharing bikes and cars, sharing portable chargers, and secondhand item transfers. In aggregation models, the information sharing platform can expand the scope of service providers for its users and obtain more user data at the same time. Super platforms mean that sharing service providers can

XSL•FO
RenderX

gain opportunities to acquire customers and increase market share at a lower cost. Users of the information sharing platforms can also gain knowledge of the quality and price of sharing services in the market and make choices more efficiently.

Although information sharing platforms such as travel platforms and car and bike sharing apps are widespread, information sharing models for older adults remain underdeveloped. Aggregation platforms can optimize users' experiences, particularly when older age people face obstacles in using smart devices. An information sharing platform suitable for older adults should help older adults filter out misinformation. On Facebook, the amount of fake news shared by users older than 65 years is almost seven times that of the youngest group [34]. According to research on the internet life of older adults in China, the proportion of middle-aged and older people who have experienced internet fraud accounts was 67.3% of the interviewed population [35]. Information sharing platforms for older adults should have the facility to block false sharing services and to provide qualified sharing service options. Information sharing platforms should also provide safe and unified payment systems to reduce worries about mobile payments and reduce fraud, which will significantly affect the older adult's use of mobile payment systems [36]. For convenient use, the design of internet products for older adults should be straightforward to use, such as the KOMP one-click computer that does not require any username and password like many other apps. KOMP's main function is to help older adults to share photos and information, and to make video calls with family or friends [37].

Although China currently does not have a sharing information platform for older adults, some telemedicine firms have begun to realize the value of such aggregating models. For example, Ping An Good Doctor, DingDang Kuaiyao, WeDoctor, and other well-known internet medical service providers in China are providing professional information searches, online diagnoses and treatment, online pharmacies, and even commercial insurance information on their platforms.

Discussion

Strategies and Recommendations

Following the Result section, we identify four challenges and make recommendations for improvements to China's sharing economy for the aging industry.

Promoting the Adaptability of Older Adults to Technology

Although China has 940 million internet users and a penetration rate of internet technology of 67% [22], the proportion of Chinese older adult internet users is far lower than that in many countries with more resources. Studies on the use of the internet by older people show that the familiarity with technology varies widely among older adults, depending on income, education, and social status [38]. Access to technology use among older adults not only makes social services more accessible but also enhances user's happiness and social connectedness [39]. The outbreak of COVID-19 has reinforced the importance of internet technology for old people by emphasizing the importance of the internet, big data, and online health care in the prevention and control of the pandemic. The application of technology in the fight against the COVID-19 virus showed the weakness of the older adult group in internet technology acquisition [40]. For example, data platforms on citizens' travel history, contacts, and disease history, which allowed citizens without COVID-19 symptoms to get a QR (Quick Response) *health code*, specifying their health status and travel rights, had been taken up by the young much more than older people. The government has identified the need to help older adults overcome obstacles to using technology [41,42], but the government, private enterprises, and community organizations should implement support to help older adults overcome obstacles in using technology, especially smartphones and the internet. Current mobile phones on the market mainly consider the needs of young people, but when designing products and providing services for older adults, enterprises should take into consideration smartphones with large screens, large fonts, long battery capacity, and simple operation. The government has been encouraging some companies to pay attention to older adult–friendly phones and list qualified products in a promotion catalog for older adults. In addition, internet service providers should consider developing models or pages suitable for older adults, providing more functions such as content reading, operation prompts, and voice assistance. Community workers could also help older adults use smart technology and the internet, for example, through community-based courses for older adults to learn to use smartphones and the internet. In a country that highly values filial piety, young people should help their elders to bridge the technological gap. At the same time, the government needs to protect the retention of traditional service methods familiar to older adults, such as retaining cash payments.

Improve the Level of Shared Services for Older Adults

There are three development problems to overcome to improve China's shared older adult care industry: insufficient development in scope, imbalance of the industrial structure, and compliance issues. Insufficient development in scope means an inadequate spread of many shared older adult care services. The *doctor sharing* model is a success story, especially during the COVID-19 pandemic, which has proved the feasibility of the people sharing model and verified both its business value and social contribution. Other shared services such as nursing attendants have failed to provide widespread resources, and there is no qualified nursing attendants information sharing platform for older adults.

Additionally, the imbalance of industrial structure reflects the differential access to shared services by region (the industrial east and coastal provinces vs underdeveloped and western regions), by urban-rural and by rich and poor older adult households. Since shared older adult market-driven service models have quasi-public characteristics, there is a need for government to engage private entrepreneurs and firms in the older adult service market to expand older adults' access to resources. Long-term inequalities by region require national government interventions. Finally, by compliance issues, we mean sharing economy providers who prioritize attributes such as the speed and scale of expansion over platform compliance

XSL•FO
RenderX

to the needs of older adults. For instance, shared nursing attendants services resulted in cases of abuse of older adults, and fundraising platforms marketing financial assistance sharing services have been subject to scams.

We recommend state interventions to address these three challenges. First, we suggest the central government should formulate guiding policies to promote the development of the shared aging care and service industry. Specifically, we recommend that the central government strengthen special research on shared aging and provide more comprehensive and specific support policies. These guiding policies should cover multiple fields such as human resources, older adult care facilities, medical services, and industry development. In addition, we recommend subnational governments at all levels should also formulate corresponding policies to promote the development of the local shared older adult care industry based on the endowment of local resources.

Second, the central government should finance an appropriate supply of shared older adult care services, especially to older adults unable to pay for shared services. Central and local governments should expand PPPs, especially in the construction of low rent apartments for the financially less well-off older adults. Changes in land use regulations should be implemented to integrate land, housing, and medical resources in the sharing model. For private enterprises participating in shared older adult care, the government should extend tax incentives.

Third, the state should improve the supervision mechanism on shared older adult care services and service providers [43]. All shared older adult care platforms should have service complaint or feedback channels to protect the rights and interests of older adults, based on industrywide standards. At present, many shared older adult care service platforms in China do not provide feedback. At the same time, these complaint and feedback channels need to be designed to suit the operating habits of older adults, like providing more voice prompts and direct telephone calls, and avoid the use of inflexible artificial intelligence responses. Local, provincial, and national regulators should exercise external supervision, improving the legal provisions to address compliance issues. Based on the framework of Figure 1, the most urgent supervisory need is to strengthen legislative supervision of crowdfunding activities controlled by for-profit companies [44].

The "One-Size-Fits-All" and the Idea of "Useless Elderly"

One problem in the provision of shared services for older adults is the lack of awareness of diversity among older-aged people. Although some older-aged people experience more chronic diseases, multimorbidities, and declining mental acuity [45], many old people reject a life of *doing nothing*. However, many old aged shared service providers view older adults as nonproductive or the *useless elderly*. For example, many apartments for older adults do not provide facilities and venues for creative activities. State regulations should require the provision of social and active centers at community-based old age care facilities and apartments, which can reduce the risk of ill health and improve happiness in old age [46,47]. Existing service providers of shared older adult care should address the spiritual needs of different types of older adult people and provide more opportunities for older adults to participate in society. According to the fourth sampling survey of the living conditions of older adults in China's urban and rural areas, 72.9% of older adults were willing to help other older adults with difficulties in the community, and 21.4% of older adults had made suggestions to the community [48]. Rather than the *useless elderly*, various large cities in China have recruited older people to volunteer for public services, such as mediating neighborhood disputes and serving as information helpers at the city's subway stations, bus stops, and tourist sites. We recommend that the government set up information sharing platforms that mainly publish city volunteer work information for older adults. Private enterprises should be encouraged to build information sharing platforms for older adult job seekers and recruitment companies so that older adults also have the opportunity to obtain labor income after retirement.

Shortage of Qualified Practitioners

An important practical reason to develop shared older adult care is the shortage of aging industry practitioners, including doctors, nurses, and nursing attendants. It is estimated that there was a shortage of 9.3 million trained nursing attendants [49], and existing doctors and nurses face high work pressure, long work hours, and low salaries, especially in public hospitals. The sharing model can facilitate the flow of information to identify high-quality health workers and to improve users' evaluation of the medical and health industry, which will allow more efficient allocation of human health care resources. Additionally, with the help of sharing platforms, health workers can serve more customers, efficiently allocate scarce human resources, and allow health practitioners to obtain higher incomes. Finally, sharing models, subject to public supervision, can promote a more standardized service to older adults, improve the working environment for health workers, and encourage new recruits to enter the health care industry.

Although the sharing economy model can help solve the problem of the shortage of health workers, government financial resources still need to be allocated to address the shortages of health workers, including doctors, nurses, and nursing attendants. First, the state should further guarantee doctor and nurse incomes, including incentive salaries for doctors and nurses. Among China's 98 industry classifications, the salary of health workers ranks 26th in the country, which is only 1.2 times the average salary in China [50]. Sharing platforms can provide incentives for doctors and nurses to earn additional income through multisite practicing and incentive payments for additional work.

Second, we recommend addressing unprotected health workers' labor rights, especially irregular labor contracts and the absence of guaranteed social insurance, where nursing attendants are especially disadvantaged. For example, the work intensity of nursing attendants is high, with a nursing attendant in a nursing home taking care of 6 to 7 older people [51], and the salary of nursing attendants are often lower than that of domestic workers. Finally, unified professional training and consistent regulations are required for professional health workers, which will help address the shortages of nurses and nursing attendants.

XSL•FO
RenderX

Conclusion

Constructing a sharing economy framework, comprising people, facilities, capital, and information, we analyzed the current status and prospects for the future development in China's sharing aging industry economy. In people sharing, we identified shortfalls in the sharing of emotional companionship, doctors, nurses, nursing attendants, and domestic helpers. Facility sharing models required improvements in the allocation of land and housing, medical devices, and other items like pensioner meals and shared medicine bins. Although crowdfunding platforms have developed fast in China, difficulties in shared capital platforms opened them to compliance issues. Information sharing is a developing field, which can optimize older users' knowledge needs and help filter out misinformation, but sharing information platforms for older adults remain both underdeveloped and unregulated.

From the analyses of the key elements in the sharing economy model, we identified four major challenges for older adults: poor adaptability to technology for older adults, mediocre quality of shared services, *one-size-fits-all* and the concept of the *useless elderly*, and the shortage of qualified practitioners. To address these gaps, we recommended specific actions by the government, communities, and enterprises. To promote the adaptability of technology for older adults, older adults need support and training to overcome obstacles in using technology, especially smartphones; enterprises should design internet products and services for older adult's convenience; and communities, especially community workers and young people, should aid older adults in smart technology and internet use. Since shared older adult market-driven service models have quasi-public characteristics, we recommend state interventions. The mindset that *one-size-fits-all* and the idea of *useless elderly* must change. The state should legislate for social and active centers at community-based old age care facilities and apartments, and the state should encourage, and enterprises provide, platforms for older adult job seekers and volunteer workers. To promote the shared health aging industry, the state should expand the number of qualified practitioners, guarantee doctor and nurse incomes, address irregular labor contracts, especially for nursing attendants, and create unified professional training to alleviate the shortages of doctors, nurses, and nursing attendants. Our analysis of China's sharing economy in the aging industry will be instructive for other countries *getting old before getting rich*.

Conflicts of Interest

None declared.

References

1. National annual data. National Bureau of Statistics. URL: https://data.stats.gov.cn/easyquery.htm?cn=C01 [accessed 2021-06-24]
2. The Ministry of Civil Affairs held a regular press conference in the fourth quarter of 2020. China Social Daily. URL: http://www.mca.gov.cn/article/xw/mtbd/202010/20201000030017.shtml [accessed 2021-06-24]
3. GDP per capita (current US$). The World Bank. URL: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD [accessed 2021-06-24]
4. Li S. How do you think "600 million people earn 1,000 yuan a month". Xinhuanet. URL: http://www.xinhuanet.com/politics/2020-06/22/c_1126144559.htm [accessed 2021-06-14]
5. Hossain M. Sharing economy: a comprehensive literature review. Int J Hospitality Manage 2020 May;87:102470. [doi: 10.1016/j.ijhm.2020.102470]
6. Acquier A, Daudigeos T, Pinkse J. Promises and paradoxes of the sharing economy: An organizing framework. Technological Forecasting Soc Change 2017 Dec;125:1-10. [doi: 10.1016/j.techfore.2017.07.006]
7. Gerwe O, Silva R. Clarifying the sharing economy: conceptualization, typology, antecedents, and effects. Acad Manage Perspect 2020 Feb;34(1):65-96. [doi: 10.5465/amp.2017.0010]
8. Palos-Sanchez PR, Correia MB. The collaborative economy based analysis of demand: study of Airbnb case in Spain and Portugal. J Theor Appl Electronic Commerce Res 2018 May;13(3):85-98. [doi: 10.4067/s0718-18762018000300105]
9. Wu H, Chan N, Zhang CJP, Ming W. The role of the sharing economy and artificial intelligence in health care: opportunities and challenges. J Med Internet Res 2019 Oct 15;21(10):e13469 [FREE Full text] [doi: 10.2196/13469] [Medline: 31617850]
10. Belk R. You are what you can access: sharing and collaborative consumption online. J Business Res 2014 Aug;67(8):1595-1600. [doi: 10.1016/j.jbusres.2013.10.001]
11. China sharing economy development report. National Information Center. URL: http://www.sic.gov.cn/News/568/10429.htm [accessed 2021-06-24]
12. The 2017 government work report was released in full. Xinhua News Agency. URL: http://www.china.com.cn/lianghui/news/2019-02/28/content_74505911.shtml [accessed 2021-06-24]
13. Speech by Xi Jinping at the Opening Ceremony of the Second World Internet Conference. Xinhuanet. URL: http://www.xinhuanet.com/politics/2015-12/16/c_1117481089.htm [accessed 2021-06-24]
14. Miller J, Ward C, Lee C, D'Ambrosio L, Coughlin J. Sharing is caring: the potential of the sharing economy to support aging in place. Gerontol Geriatr Educ 2020;41(4):407-429. [doi: 10.1080/02701960.2018.1428575] [Medline: 29364788]

XSL•FO
RenderX

15. Press release of “Research on China's Urban Home Care Services”. Ministry of Culture and Tourism of the People's Republic of China. URL: https://www.mct.gov.cn/whzx/bnsj/ltxgbj/201111/t20111128_824219.htm [accessed 2021-06-24]
16. Announcement of the Beijing Civil Affairs Bureau on the disclosure of the general survey of Beijing's home care service facilities. Beijing Civil Affairs Bureau. URL: https://www.sohu.com/a/232009364_825955 [accessed 2021-06-24]
17. The results of the fourth sample survey of the living conditions of the elderly in China's urban and rural areas are released. CNR. URL: http://old.cnr.cn/2016csy/gundong/20161010/t20161010_523186698.shtml [accessed 2021-06-24]
18. Raymo JM. Living alone in Japan: relationships with happiness and health. Demogr Res 2015;32:1267-1298 [FREE Full text] [doi: 10.4054/DemRes.2015.32.46] [Medline: 30310360]
19. Avlund K, Holstein BE, Mortensen EL, Schroll M. Active life in old age. Combining measures of functional ability and social participation. Dan Med Bull 1999 Sep;46(4):345-349. [Medline: 10514941]
20. Lavan AH, Gallagher PF, O'Mahony D. Methods to reduce prescribing errors in elderly patients with multimorbidity. Clin Interv Aging 2016;11:857-866. [doi: 10.2147/CIA.S80280] [Medline: 27382268]
21. The State Council Information Office held a press conference on the health reform and development during the 13th Five-Year Plan. State Council Information Office. URL: http://www.gov.cn/xinwen/2020-10/28/content_5555672.htm [accessed 2021-06-24]
22. China Internet Network Information Center. The 46th China Statistical Report on Internet Development. Cyberspace Administration of China. URL: http://www.cac.gov.cn/2020-09/29/c_1602939918747816.htm [accessed 2021-06-24]
23. 2019 National Hospital Internet Influence Ranking. Cyberspace Administration of China. URL: http://www.cac.gov.cn/2019-11/05/c_1574488234833118.htm [accessed 2021-06-24]
24. Schwartz TB. For fun and profit: how to install a first-rate doctor in a third-rate nursing home. N Engl J Med 1982 Mar 25;306(12):743-744. [doi: 10.1056/NEJM198203253061212] [Medline: 7062943]
25. China Statistical Yearbook 2020. National Bureau of Statistics. URL: http://www.stats.gov.cn/tjsj/ndsj/2020/indexch.htm [accessed 2021-06-24]
26. Bao S. Research on the development of China's elderly service industry. CNKI. URL: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2019&filename=1019140560.nh&v=Wbz1h0xWAuGa8OXPs%25mmd2FJrdAT0KUk4SJnZD5UTKlTPLF2HQkjf7zKXh1kZLu9sa9%25mmd2FE [accessed 2021-06-24]
27. Announcement of the General Office of the National Health Commission on the Typical Experience of the National Medical Care. General Office of the National Health Commission. URL: http://www.nhc.gov.cn/lljks/zcwj2/202009/f7c0cc74ec5c49f88b8eee67950c21eb.shtml [accessed 2021-06-24]
28. “Beijing Aging Career and Senior Service Development Report (2016-2017)” released. CNR. URL: http://old.cnr.cn/2016csy/gundong/20171101/t20171101_524008717.shtml [accessed 2021-06-24]
29. Reply of the Ministry of Civil Affairs to “Suggestion on Promoting 'Time Bank' in China”. Ministry of Civil Affairs of China. URL: http://www.mca.gov.cn/article/gk/jytabljggk/rddbjy/201810/20181000011820.shtml [accessed 2021-06-24]
30. Trial operation of Beijing's first PPP model elderly care institution. Xinhuanet. URL: http://www.xinhuanet.com/local/2017-06/29/c_1121230505.htm [accessed 2021-06-24]
31. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
32. Notice on the Reform and Innovative Pilot Program of Guangzhou Community Home Care Services. Guangzhou Home-Care Online Service Hall. URL: https://jjyl.gzmz.gov.cn/notice/toview/ec335258378e4de3b66b956d6fbd9a3b/a4688319b1f74e529d97d2dbc4b5daf8 [accessed 2021-06-24]
33. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet 2019 Sep 28;394(10204):1192-1204. [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
34. Guess A, Nagler J, Tucker J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. Sci Adv 2019 Jan;5(1):eaau4586. [doi: 10.1126/sciadv.aau4586] [Medline: 30662946]
35. Research report on the internet life of middle-aged and elderly people. Guangming Daily. URL: https://baijiahao.baidu.com/s?id=1595596223440183376&wfr=spider&for=pc [accessed 2021-06-24]
36. Liébana-Cabanillas F, Sánchez-Fernández J, Muñoz-Leiva F. Antecedents of the adoption of the new mobile payment systems: the moderating effect of age. Comput Hum Behav 2014 Jun;35:464-478. [doi: 10.1016/j.chb.2014.03.022]
37. Smith P. 15 digital solutions to support the social lives of older people. Nesta. URL: https://www.nesta.org.uk/blog/15-digital-solutions-support-social-lives-older-people/ [accessed 2021-06-24]
38. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: digital inequality among older adults. Universal Access Inf Soc 2018 May 3;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
39. Barbosa Neves B, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]

XSL•FO
RenderX

40. van Deursen AJ. Digital inequality during a pandemic: quantitative study of differences in COVID-19-related internet uses and outcomes among the general population. J Med Internet Res 2020 Aug 20;22(8):e20073 [FREE Full text] [doi: 10.2196/20073] [Medline: 32750005]
41. The General Office of the State Council issued a notice on the implementation plan to effectively solve the difficulties of the elderly in using intelligent technology. State Council. URL: http://www.gov.cn/zhengce/content/2020-11/24/content_5563804.htm [accessed 2021-06-24]
42. Ministry of Civil Affairs: strive to solve the smart technology difficulties encountered by the elderly in civil affairs services. State Council. URL: http://www.gov.cn/xinwen/2020-10/24/content_5553807.htm [accessed 2021-06-24]
43. Ingenhoff D, Koelling AM. The potential of web sites as a relationship building tool for charitable fundraising NPOs. Public Relations Rev 2009 Mar;35(1):66-73. [doi: 10.1016/j.pubrev.2008.09.023]
44. Charity Law of the People's Republic of China. China Law. URL: http://search.chinalaw.gov.cn/law/searchTitleDetail?LawID=332773&Query=%E6%85%88%E5%96%84%E6%B3%95&IsExact=&PageIndex=1 [accessed 2021-06-24]
45. Wu Y. Establish a positive outlook on aging and promote the healthy and sustainable development of the elderly care service industry in the new era. Chinese Social Work. URL: https://tinyurl.com/2jpwpsrv [accessed 2021-06-24]
46. Graney MJ. Happiness and social participation in aging. J Gerontol 1975 Nov;30(6):701-706. [doi: 10.1093/geronj/30.6.701] [Medline: 1184930]
47. Young FW, Glasgow N. Voluntary social participation and health. Res Aging 2016 Aug 18;20(3):339-362. [doi: 10.1177/0164027598203004]
48. The results of the fourth sample survey of the living conditions of the elderly in China's urban and rural areas are released. CNR. URL: http://old.cnr.cn/2016csy/gundong/20161010/t20161010_523186698.shtml [accessed 2021-06-24]
49. Zou M. How to fill the gap for senior care professionals. Finance China. URL: http://finance.china.com.cn/roll/20160706/3799261.shtml [accessed 2021-06-24]
50. Wang H. Research on the salary incentive mechanism and model of doctors in public hospitals. CNKI. URL: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2021&filename=1020412093.nh&v=%25mmd2FYUNE9RRuVasq%25mmd2FTt4oG5I9tRJBtd51V1yhjzTe07zcLZC%25mmd2FNnSvf9DrYF%25mmd2FGoRv903 [accessed 2021-06-24]
51. Xu S. Research on the construction of nursing staff team in the elderly care institutions with integrated medical care in Wuhan. CNKI. URL: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202001&filename=1019907810.nh&v=%25mmd2BJQnHclazFZDTS%25mmd2BtlvCHusQJqE0Z1Os5Q6qZnHOolLrED0Lj%25mmd2FVuIq5wsymT%25mmd2F1guv [accessed 2021-06-24]

Abbreviations

GDP: gross domestic product
O2O: online-to-offline
PPP: private-public partnership
QR: Quick Response

Edited by R Kukafka; submitted 05.02.21; peer-reviewed by W Zhang, P Palos-Sanchez; comments to author 30.03.21; revised version received 03.04.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Hu Y, Wang J, Nicholas S, Maitland E
The Sharing Economy in China's Aging Industry: Applications, Challenges, and Recommendations
J Med Internet Res 2021;23(7):e27758
URL: https://www.jmir.org/2021/7/e27758
doi:10.2196/27758
PMID:34255691

©Yaolin Hu, Jian Wang, Stephen Nicholas, Elizabeth Maitland. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Recursive Partitioning vs Computerized Adaptive Testing to Reduce the Burden of Health Assessments in Cleft Lip and/or Palate: Comparative Simulation Study

Conrad J Harrison[1], BSc, MBBS, MRCS; Chris J Sidey-Gibbons[2], PhD; Anne F Klassen[3], DPhil; Karen W Y Wong Riff[4], MD, PhD; Dominic Furniss[1], DM; Marc C Swan[5], DPhil; Jeremy N Rodrigues[6,7], MBA, PhD

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[2]MD Anderson Center for INSPiRED Cancer Care, University of Texas, Houston, TX, United States

[3]Department of Pediatrics, McMaster University, Hamilton, ON, Canada

[4]Department of Plastic and Reconstructive Surgery, Hospital for Sick Children, Toronto, ON, Canada

[5]Spires Cleft Centre, John Radcliffe Hospital, Oxford, United Kingdom

[6]Warwick Clinical Trials Unit, University of Warwick, Coventry, United Kingdom

[7]Department of Plastic Surgery, Stoke Mandeville Hospital, Buckinghamshire Healthcare NHS Trust, Aylesbury, United Kingdom

Corresponding Author:
Conrad J Harrison, BSc, MBBS, MRCS
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford
The Botnar Research Centre
Old Road
Oxford, OX3 7LD
United Kingdom
Phone: 44 1865 227 374
Email: conrad.harrison@medsci.ox.ac.uk

Abstract

Background: Computerized adaptive testing (CAT) has been shown to deliver short, accurate, and personalized versions of the CLEFT-Q patient-reported outcome measure for children and young adults born with a cleft lip and/or palate. Decision trees may integrate clinician-reported data (eg, age, gender, cleft type, and planned treatments) to make these assessments even shorter and more accurate.

Objective: We aimed to create decision tree models incorporating clinician-reported data into adaptive CLEFT-Q assessments and compare their accuracy to traditional CAT models.

Methods: We used relevant clinician-reported data and patient-reported item responses from the CLEFT-Q field test to train and test decision tree models using recursive partitioning. We compared the prediction accuracy of decision trees to CAT assessments of similar length. Participant scores from the full-length questionnaire were used as ground truth. Accuracy was assessed through Pearson's correlation coefficient of predicted and ground truth scores, mean absolute error, root mean squared error, and a two-tailed Wilcoxon signed-rank test comparing squared error.

Results: Decision trees demonstrated poorer accuracy than CAT comparators and generally made data splits based on item responses rather than clinician-reported data.

Conclusions: When predicting CLEFT-Q scores, individual item responses are generally more informative than clinician-reported data. Decision trees that make binary splits are at risk of underfitting polytomous patient-reported outcome measure data and demonstrated poorer performance than CATs in this study.

(J Med Internet Res 2021;23(7):e26412) doi:10.2196/26412

KEYWORDS

cleft Lip; cleft palate; patient-reported outcome measures; outcome assessment; CLEFT-Q; computerized adaptive test; machine learning; decision tree; regression tree

XSL•FO
RenderX

Introduction

Computerized Adaptive Testing

Computerized adaptive testing (CAT) describes the use of algorithms to shorten and personalize questionnaires by selectively administering only the most relevant items to an individual based on the responses they have already given during that assessment [1]. The item that is expected to be most relevant (ie, most informative) in a general population is picked first [2,3]. Once the individual has responded, an algorithm predicts the person's score and selects the most relevant item to ask next, based on the predicted score and an item selection criterion. This continues iteratively until a stopping rule is met.

There are some limitations in this approach to personalized assessments. Firstly, this method is only possible in patient-reported outcome measures (PROMs) that fit item response theory (IRT) or Rasch models, and not those that only meet the structural requirements of classical test theory [1,4]. Secondly, an individual's score must be calculated after each response to select the next item, which can be computationally demanding in situations where the assessment has many different items to choose from (ie, item-banking) and may cause a time delay between each item. Thirdly, CAT assessments do not directly incorporate clinician-reported variables. These variables, which can be automatically captured from a person's electronic health record, may be very informative and can potentially improve the efficiency and accuracy of personalized assessments.

Recursive Partitioning

Recursive partitioning is a form of machine learning that involves iteratively splitting labeled data sets into subgroups to minimize the within-subgroup variance of an outcome, such as a PROM score [5]. Recent studies have explored the use of personalized health assessments based on decision trees constructed with similar techniques [6-8]. These trees split respondents into subgroups based on their responses to individual items.

The use of decision trees in personalized health assessment may be appealing because they are not restricted by IRT model requirements, and trees are developed *a priori* (ie, they do not need to calculate a person's score between each item), attenuating potential lag time [9]. It may also be possible to use recursive partitioning to split data based on clinical variables other than item responses, meaning that, unlike traditional CAT assessments, decision trees could automatically incorporate clinical information known to predict a person's score. For example, the frequency of inhaler prescriptions could guide item selection in an asthma assessment, and step count could be incorporated into an assessment of mobility. However, many PROMs comprise polytomous response options, which may not be handled well by binary decision nodes.

To our knowledge, the potential for recursive partitioning to improve the accuracy and efficiency of CAT by incorporating clinician-reported variables into patient-reported assessments has not been explored prior to this study.

The CLEFT-Q

A cleft lip and/or palate (a split in the upper lip, gum, and/or roof of the mouth) is one of the most common birth anomalies. It affects 1 in 700 births and can impact various health domains, including appearance, speech, and psychosocial development [10].

In this study, we developed CAT models and decision trees for scales of the CLEFT-Q. The CLEFT-Q is a PROM intended for use in children and young adults born with a cleft lip and/or palate [11]. It includes 12 independent scales that have met Rasch model requirements, measuring the perceived appearance of the respondent's face, nose, nostrils, lips, cleft lip scar, jaw, and teeth, as part of an "appearance" domain; speech function, as part of a "facial function" domain; and speech distress, psychological function, school function, and social function, as part of a "quality of life" domain. Differences in CLEFT-Q scale scores have been associated with the overall patient-reported perception of appearance [11] and several clinical variables, including cleft type [12], use of psychological therapy, [11] and clinician-reported plans for future surgery [13].

Hypothesis

We tested the null hypothesis that adaptive assessments incorporating clinical variables and item responses (created using recursive partitioning) would predict CLEFT-Q scale scores with similar accuracy to CAT assessments of a comparable length.

Methods

Software

We conducted our analysis using R (version 4.0.0) with the following packages: dplyr (version 1.0.0), foreign (version 0.8-80), mirt (version 1.32.1), mirtCAT (version 1.9.3), partykit (version 1.2-8), rpart (version 4.1-15), rpart.plot (version 3.0.8), stringr (version 1.4.0) and ggplot2 (version 3.3.2).

Study Participants

We used data from the CLEFT-Q field test to construct and test our models. This prospective study recruited 2434 participants across 12 countries born with a cleft lip and/or palate, aged 8 to 29 years, from October 2014 to November 2016. Responses to all items in relevant full-length CLEFT-Q scales were collected in addition to clinical information. The CLEFT-Q field test study and its participants have been described in detail elsewhere [11,14].

These participants' complete response sets were used to develop and test CAT and decision tree models. Participants with a cleft palate only were excluded from analyses relating to the lip and cleft lip scar scales. Patients without a cleft palate were excluded from analyses relating to the speech function and speech distress scales. This reflects populations expected to use these assessments in clinical practice.

Missing Data

Some functions in the mirt package require complete response sets [15]. Where applicable, we excluded participants with

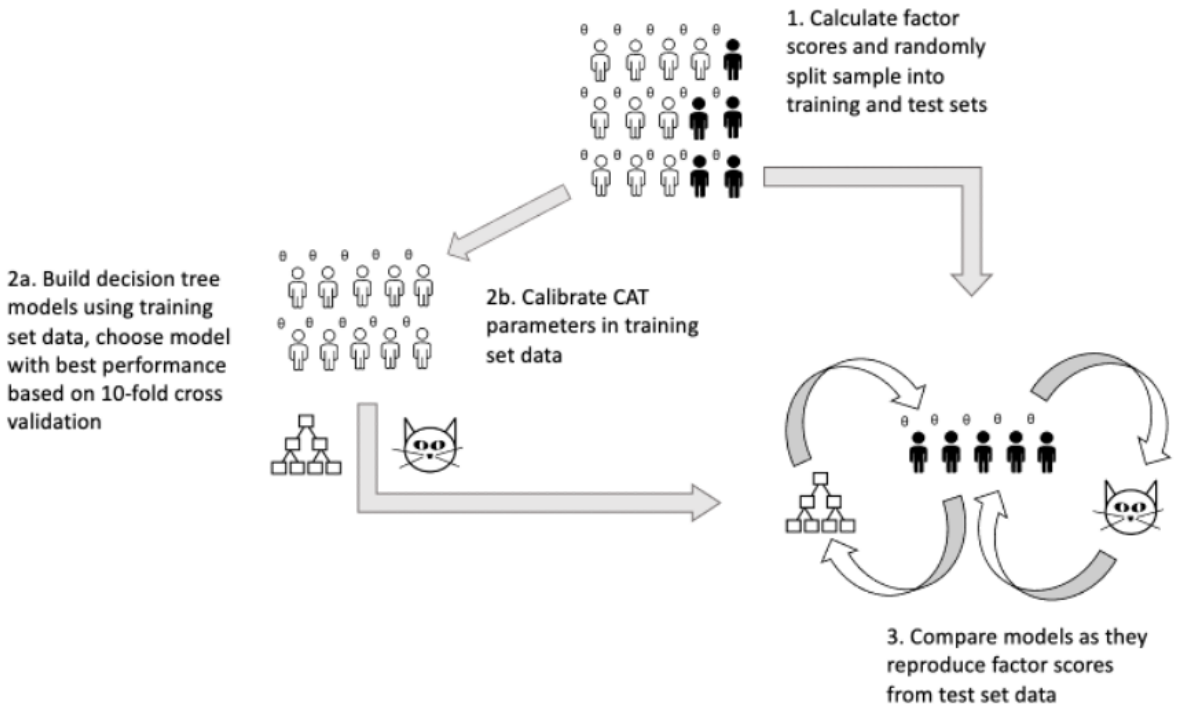
XSL•FO
RenderX

incomplete response sets listwise. Otherwise, missing data were handled with pairwise exclusion. In the CLEFT-Q field test, only participants aged 12 to 29 years completed the jaw scale, only participants born with a cleft lip completed the cleft lip scar scale, and only participants attending school completed the school scale [11]. There were no other apparent patterns in missing data. We provide descriptive statistics, including a summary of missing data in a demographic table (see Multimedia Appendix 1).

Ground Truth

Rasch models were developed for each scale using all item response sets from the whole sample. Model parameters were estimated using a fixed quadrature expectation-maximization (EM) approach [15]. For each participant, expected *a posteriori* (EAP) factor scores were calculated for each scale using full-length scale responses and the respective Rasch model parameters. These scores are presented as person-location logits and represent ground truth in these experiments (Figure 1).

Figure 1. Schematic describing the steps of the CAT vs recursive partitioning experiment. CAT: computerized adaptive test.



Data Splitting

Participants with complete response sets were selected from the study sample and randomly split into training and test sets in a 2:1 ratio for each scale. Next, training samples were used to create new Rasch models using the EM algorithm, and CAT models were generated using the parameters of these training sample Rasch models. Finally, decision trees were built through recursive partitioning in 10-fold cross-validations of the same training samples.

Sample Size

All of our training data sets approached or exceeded a sample size of 1000 (see Multimedia Appendix 1), which is likely to be sufficient for accurate regression tree development based on previous simulation studies [16]. For CAT, sample sizes of 250 or more can provide definitive Rasch model calibrations with over 99% confidence [17].

Item Responses

The data sets used to train our decision tree models included responses to all items in each CLEFT-Q scale. Items in the speech distress and speech function scale had 3 response options, and all other items had 4 response options. In addition to these items, the CLEFT-Q field test included 7 questions that asked respondents about the overall appearance of their face, nose, nostrils, lips, cleft lip scar, teeth, and jaws. There were 4 response options to these items. These responses were included in the relevant decision tree training data set, meaning, for example, that the decision tree nose assessment was able to administer an additional item that asked about overall nose appearance. The CAT model could not use this item as it was not part of the Rasch-validated scale.

Clinical Variables

In addition to item responses, the training data sets included several clinically relevant variables, for example, age, gender, cleft type, planned future treatments, and patient-perceived appearance scores. The variables included in each model are listed in Multimedia Appendix 2.

Decision Tree Training

Regression trees could make binary splits based on item responses or any of the included clinical variables. We grew trees to a prespecified maximum depth corresponding to the number of items permitted by their CAT comparator. Tree

XSL•FO
RenderX

growth was not limited by observations per node or by a prespecified complexity parameter. Branches that did not improve model fit were removed retrospectively through cost-complexity pruning. For this, we used training data to choose complexity parameters that minimized prediction error in a 10-fold cross-validation.

In other words, for each scale, many different decision trees were created from and evaluated in different folds of the training data set. The model that predicted scores with the least error in the training data set cross-validation was selected for final assessment in the test data set.

Computerized Adaptive Tests

The CATs were programmed to select items using a minimum expected posterior variance item selection criterion [18]. Finally, participants were scored using an EAP approach [15].

Assessment Length

We aimed to compare the accuracy of CAT assessments and decision trees that had a similar number of items. For this reason, we chose fixed-length stopping rules to limit the number of items administered in the CAT models. The CAT algorithms would continue to administer items until the fixed-length stopping rule was met, and they were not limited by time, standard error of measurement, or any other stopping criterion. It was not possible to predetermine decision tree assessment lengths. This is because splits could be made based on clinical variables, and some branches were pruned, creating inconsistent assessment lengths. Instead, tree growth was limited by depth, and the number of items required to reach each terminal node was recorded. If a decision tree made a split based on overall patient-reported appearance (ie, the additional question posed to CLEFT-Q field test participants, not included in the scale), we counted this as an item. Splits based on clinician-reported variables were not counted as items. For each scale, we compared models at 2 maximum assessment lengths, which were approximately 75% and 50% of the length of the whole scale.

Comparison Methods

For each respondent in the test data set, person-location logits were predicted by decision trees and their CAT comparators in Monte Carlo simulations. For each model, we recorded the mean number of items administered, which items and clinical variables were used to make predictions, the Pearson's correlation coefficient of predictions and ground truth, and the mean absolute error (MAE) of the predictions with respect to ground truth. Additionally, we calculated the root mean squared error of predictions, which is typically reported in CAT simulations and tends to penalize large errors that are potentially important in this context [19].

Squared CAT errors and squared decision tree errors were compared using a two-sided Wilcoxon signed-rank test.

Results

Assessment length was generally similar between decision trees and CAT assessments (see Multimedia Appendix 1). Notable exceptions to this were the nose assessments limited to 9 items (mean of 7.32 items in decision tree assessments vs 9.00 items in CAT assessments), and the speech function assessments were limited to 9 items (mean of 7.01 items in decision tree assessments vs 9.00 items in CAT assessments).

For most comparisons, the squared error was significantly higher ($P<.001$) in decision tree predictions than in predictions made by CAT assessments at comparable or slightly shorter mean assessment lengths. The poor accuracy of the decision tree models compared to CAT was also captured by correlation coefficients and MAE values (see Multimedia Appendix 1).

While information about age, gender, cleft type, laterality, and patient-reported overall appearance scores were used by some of the deeper decision trees, these algorithms tended to make splits based on CLEFT-Q scale item responses preferentially (ie, other variables were either not used, used at deeper levels, or removed through pruning). An exception to this was the nose assessment decision tree, which made its first split based on the patient-reported overall appearance of the nose.

Discussion

Principal Findings

This study has shown that it is technically possible for decision trees built through recursive partitioning to use clinician-reported data to reduce the patient-reported assessment burden in a range of cleft-related health domains. However, this approach has demonstrated little clinical value with regard to the CLEFT-Q. Decision trees preferentially made splits based on patient-reported item responses and not clinician-reported data. One way to interpret this in real-world terms is that the clinical variables used in this study are less important than almost any individual CLEFT-Q item response for measuring a person's cleft-related health state (Figure 2). However, this finding is not necessarily generalizable. Differences in cleft phenotype may have a relatively mild impact on health constructs measured by the CLEFT-Q. Clinician-reported variables may be more salient in other health domains. For example, "history of spinal cord injury" may be a powerful predictor of the physical functioning score.

Our second finding was that decision trees produced more error than CAT assessments of a comparable length (Figure 3; Multimedia Appendix 1). CLEFT-Q scales are short (6 to 12 items), which means CAT item selection is relatively computationally undemanding in this specific case. Any benefits in lag time gained by the recursive partitioning approach are unlikely to outweigh this loss of accuracy. Therefore, we would not advocate the use of binary decision trees for adaptive CLEFT-Q assessments.

In this study, we used recursive partitioning for tree construction, which is limited to binary data splits (ie, individuals can be categorized into only 2 subgroups after providing an item response; Figure 2). There are 3 or 4 possible responses to each CLEFT-Q item, and therefore the CAT models can categorize a respondent in 3 or 4 ways each time they provide an item response. For polytomous PROM scales, there are far fewer attainable scores in a binary decision tree than in a CAT assessment of equivalent length. For example, in the speech

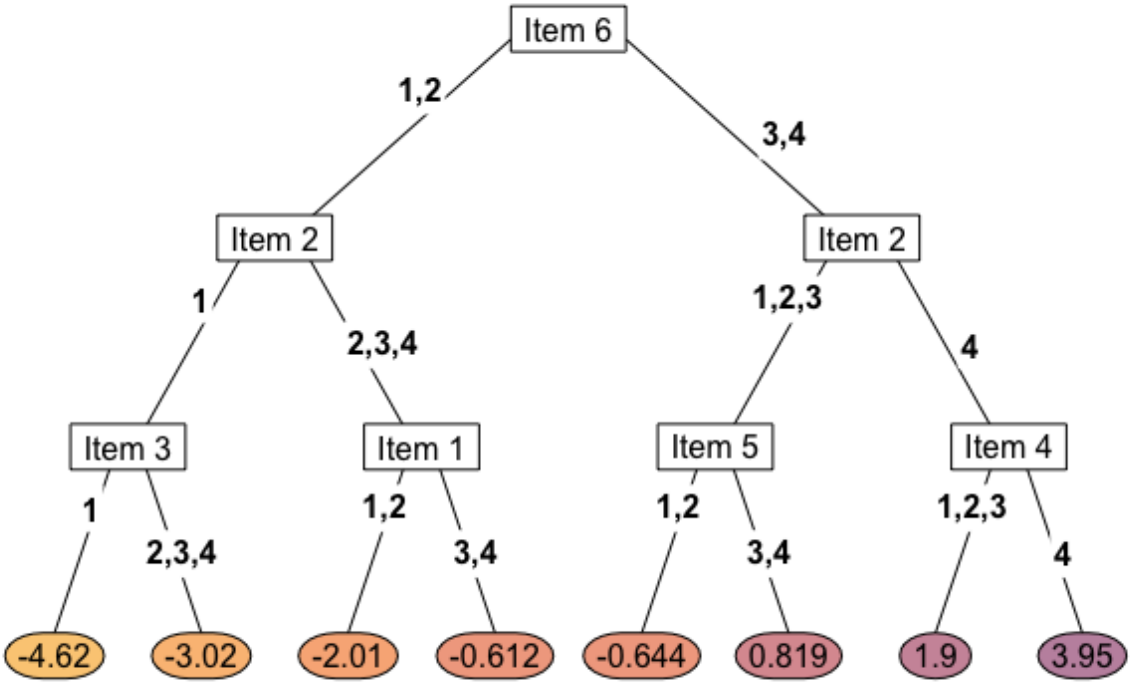
XSL•FO
RenderX

function assessment at a 6-item limit, there were 85 unique scores achieved by our CAT models compared to 45 in the decision tree group. Our decision trees, therefore, underfit the test data to an extent.

A way to overcome this in future work could be through constructing trees with nonbinary splits, for example, by treating all variables as categorical and using chi-square automatic interaction detection (CHAID) [20,21]. This technique could have a higher risk of overfitting, although this might be mitigated through split rules and by merging similar branches [6]. Future work is needed to test whether CHAID could create more efficient, more accurate, adaptive assessments by incorporating non-PROM data at no extra burden to the assessment taker.

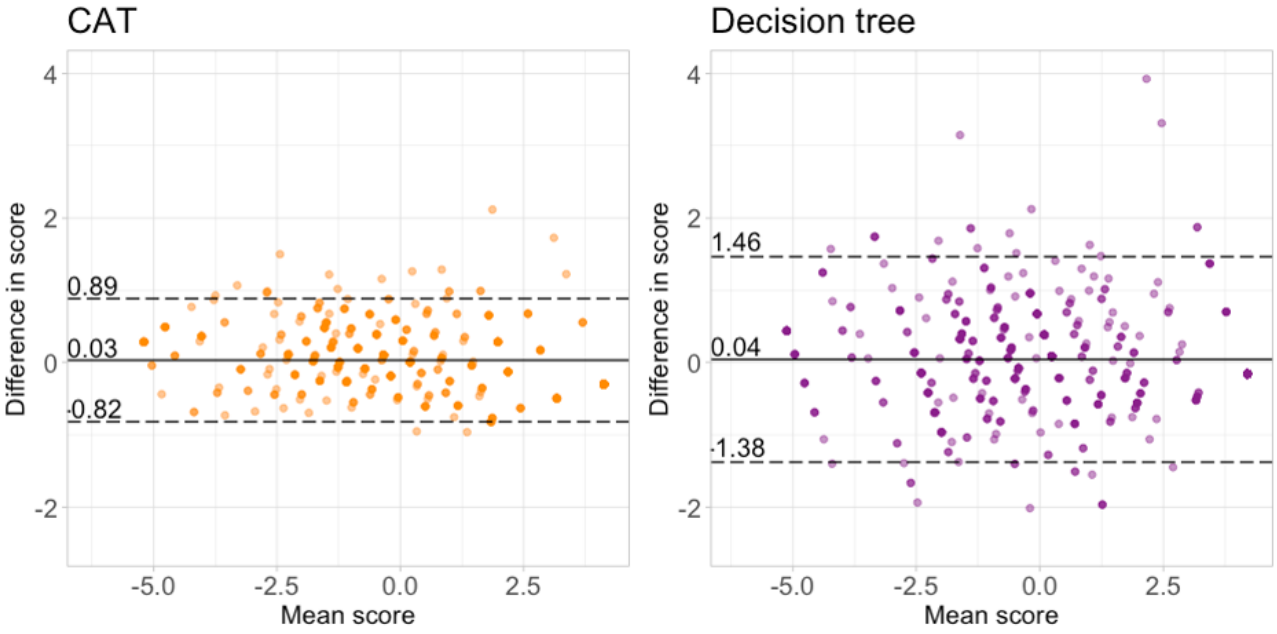
A limitation of this study is that some comparisons were made on assessments of unequal mean length; for example, the nose assessment was limited to 9 items, and the speech function assessment was limited to 9 items. In these cases, it is difficult to conclude the relative accuracy of CAT and decision trees, as these models used different quantities of patient-reported information to make their measurements. In addition, decision trees tended to pose fewer questions than their CAT comparators due to the pruning process and the use of clinical variables. However, even with this limitation in mind, our findings support the use of CAT over binary trees for personalized CLEFT-Q assessments in clinical practice.

Figure 2. Dendrogram of the three-item decision tree for the nostrils scale. Nodes are represented by rectangular boxes. Branches are labelled with splitting criteria . Leaves (terminal nodes) are represented by coloured ovals and are labelled with the mean person location logit for training dataset respondents falling into that subgroup.



XSL•FO
RenderX

Figure 3. Bland-Altman plots comparing the accuracy of personalized CLEFT-Q lip appearance assessments. The left panel demonstrates the results for a computerized adaptive test with a mean length of 5.00 items. The right panel demonstrates the results for a decision tree with a mean length of 4.95 items.



Conclusions

Even with knowledge of clinician-reported variables, the decision tree models described in this study achieve less accurate CLEFT-Q score estimates than CATs of similar length. Decision trees with binary splits are at risk of underfitting polytomous PROM scale data. Future work could focus on the application of decision trees with nonbinary splits to this context.

Acknowledgments

For this research project, CH is funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship (grant NIHR300684). JR is funded by an NIHR postdoctoral fellowship (PDF-2017-10-075). The NIHR Oxford Biomedical Research Centre supports DF. The CLEFT-Q field-test study was funded by the Canadian Institute of Health Research (CIHR; grant FRN-130515).

This article presents independent research funded by the NIHR and the CIHR. The views expressed are those of the authors and not necessarily those of the CIHR, NIHR, the National Health Service, or the Department of Health and Social Care.

Conflicts of Interest

The CLEFT-Q is owned by McMaster University and the Hospital for Sick Children. AK and KWR are codevelopers of the CLEFT-Q and, as such, could potentially receive a share of any license revenues based on their institution's inventor sharing policy. The other authors have no conflicts of interest to declare related to the content of this article.

Multimedia Appendix 1
Supplement 1 - results.
[XLSX File (Microsoft Excel File), 22 KB - jmir_v23i7e26412_app1.xlsx]

Multimedia Appendix 2
Supplement 2.
[DOCX File , 13 KB - jmir_v23i7e26412_app2.docx]

References

1. Weiss DJ, Vale CD. Adaptive Testing. Applied Psychology 1987 Jul;36(3-4):249-262. [doi: 10.1111/j.1464-0597.1987.tb01190.x]
2. Cai L, Choi K, Hansen M, Harrell L. Item response theory. Annu. Rev. Stat. Appl 2016 Jun;3(1):297-321. [doi: 10.1146/annurev-statistics-041715-033702]

XSL•FO
RenderX

3. Nguyen TH, Han H, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient 2014 Jan 9;7(1):23-35 [FREE Full text] [doi: 10.1007/s40271-013-0041-0] [Medline: 24403095]
4. Weiss DJ. Better data from better measurements using computerized adaptive testing. JMMSS 2011 Sep 23;6(1):1-27. [doi: 10.2458/azu_jmmss.v2i1.12351]
5. Atkinson E, Therneau T. An introduction to recursive partitioning using the RPART routines. CRAN. 2019 Apr 11. URL: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf [accessed 2021-07-15]
6. Rodríguez-Cuadrado J, Delgado-Gómez D, Laria JC, Rodríguez-Cuadrado S. Merged tree-CAT: a fast method for building precise computerized adaptive tests based on decision trees. Expert Systems with Applications 2020 Apr;143:113066-113066. [doi: 10.1016/j.eswa.2019.113066]
7. Michel P, Baumstarck K, Loundou A, Ghattas B, Auquier P, Boyer L. Computerized adaptive testing with decision regression trees: an alternative to item response theory for quality of life measurement in multiple sclerosis. PPA 2018 Jun;Volume 12:1043-1053. [doi: 10.2147/ppa.s162206]
8. Delgado-Gomez D, Baca-Garcia E, Aguado D, Courtet P, Lopez-Castroman J. Computerized adaptive test vs. decision trees: development of a support decision system to identify suicidal behavior. J Affect Disord 2016 Dec;206:204-209. [doi: 10.1016/j.jad.2016.07.032] [Medline: 27475891]
9. Delgado-Gómez D, Laria JC, Ruiz-Hernández D. Computerized adaptive test and decision trees: a unifying approach. Expert Systems with Applications 2019 Mar;117:358-366. [doi: 10.1016/j.eswa.2018.09.052]
10. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. The Lancet 2009 Nov;374(9703):1773-1785. [doi: 10.1016/s0140-6736(09)60695-4]
11. Klassen AF, Riff KWW, Longmire NM, Albert A, Allen GC, Aydin MA, et al. Psychometric findings and normative values for the CLEFT-Q based on 2434 children and young adult patients with cleft lip and/or palate from 12 countries. CMAJ 2018 Apr 16;190(15):E455-E462 [FREE Full text] [doi: 10.1503/cmaj.170289] [Medline: 29661814]
12. Wong Riff KWY, Tsangaris E, Forrest CR, Goodacre T, Longmire NM, Allen G, et al. CLEFT-Q. Plastic and Reconstructive Surgery 2019;144(1):78e-88e. [doi: 10.1097/prs.0000000000005723]
13. Harrison CJ, Rae C, Tsangaris E, Wong Riff KW, Swan MC, Goodacre TE, et al. Further construct validation of the CLEFT-Q: ability to detect differences in outcome for four cleft-specific surgeries. J Plast Reconstr Aesthet Surg 2019 Dec;72(12):2049-2055. [doi: 10.1016/j.bjps.2019.07.029] [Medline: 31488380]
14. Wong Riff KWY, Tsangaris E, Goodacre T, Forrest CR, Pusic AL, Cano SJ, et al. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). BMJ Open 2017 Jan 11;7(1):e015467 [FREE Full text] [doi: 10.1136/bmjopen-2016-015467] [Medline: 28077415]
15. Chalmers RP. mirt : a multidimensional item response theory package for the R environment. J. Stat. Soft 2012;48(6):1-29. [doi: 10.18637/jss.v048.i06]
16. Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S. Decision trees in epidemiological research. Emerg Themes Epidemiol 2017 Sep 20:14-11. [doi: 10.1186/s12982-017-0064-4]
17. Linacre J. Sample size and item calibration stability. Rasch Measurement Transactions. URL: https://www.rasch.org/rmt/rmt74m.htm [accessed 2021-07-15]
18. Chalmers RP. Generating adaptive and non-Adaptive test interfaces for multidimensional item Response Theory Applications. J. Stat. Soft 2016;71(5):1-38. [doi: 10.18637/jss.v071.i05]
19. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci. Model Dev 2014 Jun 30;7(3):1247-1250. [doi: 10.5194/gmd-7-1247-2014]
20. van der Oest MJ, Porsius JT, MacDermid JC, Slijper HP, Selles RW. Item reduction of the patient-rated wrist evaluation using decision tree modelling. Disability and Rehabilitation 2019 Feb 10;42(19):2758-2765. [doi: 10.1080/09638288.2019.1566407]
21. Jansen MC, van der Oest MJ, Slijper HP, Porsius JT, Selles RW. Item reduction of the boston carpal tunnel questionnaire using decision tree modeling. Archives of Physical Medicine and Rehabilitation 2019 Dec;100(12):2308-2313. [doi: 10.1016/j.apmr.2019.04.021]

Abbreviations

CAT: computerized adaptive test
CHAID: chi-square automatic interaction detection
EAP: expected a posteriori
EM: expectation-maximization
IRT: item response theory
MAE: mean absolute error
PROM: patient-reported outcome measure

XSL•FO
RenderX

Edited by R Kukafka; submitted 10.12.20; peer-reviewed by J Chaplin, A Masino; comments to author 30.01.21; revised version received 25.02.21; accepted 24.05.21; published 30.07.21.

Please cite as:
Harrison CJ, Sidey-Gibbons CJ, Klassen AF, Wong Riff KWY, Furniss D, Swan MC, Rodrigues JN
Recursive Partitioning vs Computerized Adaptive Testing to Reduce the Burden of Health Assessments in Cleft Lip and/or Palate: Comparative Simulation Study
J Med Internet Res 2021;23(7):e26412
URL: https://www.jmir.org/2021/7/e26412
doi:10.2196/26412
PMID:34328443

©Conrad J Harrison, Chris J Sidey-Gibbons, Anne F Klassen, Karen W Y Wong Riff, Dominic Furniss, Marc C Swan, Jeremy N Rodrigues. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study

Stanislav Nikolov[1*], MEng; Sam Blackwell[1*], MEng; Alexei Zverovitch[2*], PhD; Ruheena Mendes[3], MB ChB; Michelle Livne[2], PhD; Jeffrey De Fauw[1], BSc; Yojan Patel[2], BA; Clemens Meyer[1], MSc; Harry Askham[1], MSc; Bernadino Romera-Paredes[1], PhD; Christopher Kelly[2], PhD; Alan Karthikesalingam[2], PhD; Carlton Chu[1], PhD; Dawn Carnell[3], MD; Cheng Boon[4], MB ChB; Derek D'Souza[3], MSc; Syed Ali Moinuddin[3], MSc; Bethany Garie[1], BMRSc (RT); Yasmin McQuinlan[1], BRT; Sarah Ireland[1], BMRSc (RT); Kiarna Hampton[1], MPH; Krystle Fuller[1], BAppSc (RT); Hugh Montgomery[5], BSc, MB BS, MD; Geraint Rees[5], PhD; Mustafa Suleyman[6]; Trevor Back[1], PhD; Cían Owen Hughes[2*], MBChB, MRCS, MSc; Joseph R Ledsam[7*], MB ChB; Olaf Ronneberger[1*], PhD

[1]DeepMind, London, United Kingdom

[2]Google Health, London, United Kingdom

[3]University College London Hospitals NHS Foundation Trust, London, United Kingdom

[4]Clatterbridge Cancer Centre NHS Foundation Trust, Liverpool, United Kingdom

[5]University College London, London, United Kingdom

[6]Google, London, United Kingdom

[7]Google AI, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Cían Owen Hughes, MBChB, MRCS, MSc
Google Health
6 Pancras Square
London, N1C 4AG
United Kingdom
Phone: 1 650 253 0000
Fax: 1 650 253 0001
Email: cianh@google.com

Abstract

Background: Over half a million individuals are diagnosed with head and neck cancer each year globally. Radiotherapy is an important curative treatment for this disease, but it requires manual time to delineate radiosensitive organs at risk. This planning process can delay treatment while also introducing interoperator variability, resulting in downstream radiation dose differences. Although auto-segmentation algorithms offer a potentially time-saving solution, the challenges in defining, quantifying, and achieving expert performance remain.

Objective: Adopting a deep learning approach, we aim to demonstrate a 3D U-Net architecture that achieves expert-level performance in delineating 21 distinct head and neck organs at risk commonly segmented in clinical practice.

Methods: The model was trained on a data set of 663 deidentified computed tomography scans acquired in routine clinical practice and with both segmentations taken from clinical practice and segmentations created by experienced radiographers as part of this research, all in accordance with consensus organ at risk definitions.

Results: We demonstrated the model's clinical applicability by assessing its performance on a test set of 21 computed tomography scans from clinical practice, each with 21 organs at risk segmented by 2 independent experts. We also introduced surface Dice similarity coefficient, a new metric for the comparison of organ delineation, to quantify the deviation between organ at risk surface contours rather than volumes, better reflecting the clinical task of correcting errors in automated organ segmentations. The model's generalizability was then demonstrated on 2 distinct open-source data sets, reflecting different centers and countries to model training.

XSL•FO
RenderX

Conclusions: Deep learning is an effective and clinically applicable technique for the segmentation of the head and neck anatomy for radiotherapy. With appropriate validation studies and regulatory approvals, this system could improve the efficiency, consistency, and safety of radiotherapy pathways.

(J Med Internet Res 2021;23(7):e26151) doi:10.2196/26151

KEYWORDS

radiotherapy; segmentation; contouring; machine learning; artificial intelligence; UNet; convolutional neural networks; surface DSC

Introduction

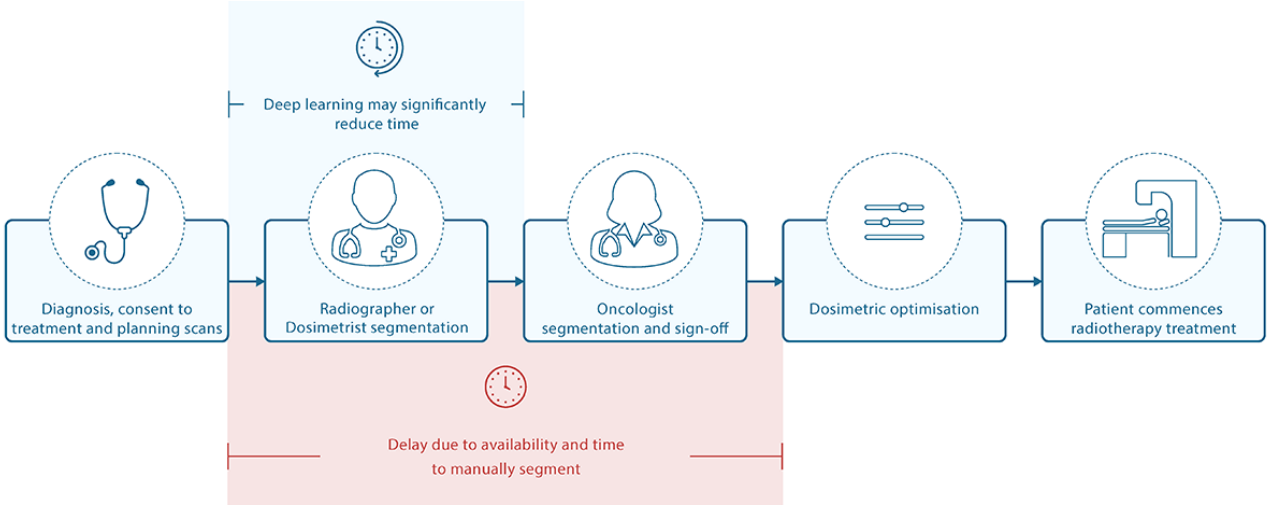
Background

Each year, 550,000 people worldwide are diagnosed with cancer of the head and neck [1]. This incidence is rising [2] and more than doubling in certain subgroups over the last 30 years [3-5]. Where available, most patients will be treated with radiotherapy, which targets the tumor mass and areas at high risk of microscopic tumor spread. However, strategies are needed to mitigate the dose-dependent adverse effects that result from incidental irradiation of normal anatomical structures (*organs at risk*) [6-9].

Thus, the efficacy and safety of head and neck radiotherapy depends on the accurate delineation of organs at risk and tumors, a process known as segmentation or contouring. However, the fact that this process is predominantly done manually means that results may be both inconsistent and imperfectly accurate [10], leading to large inter- and intrapractitioner variability even among experts and thus variation in care quality [11].

Segmentation is also very time consuming: an expert can spend 4 hours or more on a single case [12]. The duration of resulting delays in treatment initiation (Figure 1) is associated with an increased risk of both local recurrence and overall mortality [13-15]. Increasing demands for, and shortages of, trained staff already place a heavy burden on health care systems, which can lead to long delays for patients as radiotherapy is planned [16,17], and the continued rise in head and neck cancer incidence may make it impossible to maintain even current temporal reporting standards [4]. Such issues also represent a barrier to *adaptive radiotherapy*—the process of repeated scanning, segmentation, and radiotherapy planning throughout treatment, which maintains the precision of tumor targeting (and organ at risk avoidance) in the face of treatment-related anatomic changes such as tumor shrinkage [18].

Figure 1. A typical clinical pathway for radiotherapy. After a patient is diagnosed and the decision is made to treat with radiotherapy, a defined workflow aims to provide treatment that is both safe and effective. In the United Kingdom, the time delay between decision to treat and treatment delivery should be no greater than 31 days. Time-intensive manual segmentation and dose optimization steps can introduce delays to treatment.



Automated (ie, computer-performed) segmentation has the potential to address these challenges. However, most segmentation algorithms in clinical use are atlas based, producing segmentations by fitting previously labeled reference images to the new target scan. This might not sufficiently account for either postsurgical changes or the variability in normal anatomical structures that exist between patients, particularly when considering the variable effect that tumors may have on local anatomy; thus, they may be prone to systematic error. To date, such algorithm-derived segmentations still require significant manual editing, perform at expert levels on only a small number of organs, demonstrate an overall performance in clinical practice inferior to that of human experts, and have failed to significantly improve clinical workflows [19-26].

In recent years, deep learning–based algorithms have proven capable of delivering substantially better performance than traditional segmentation algorithms. Several deep learning–based approaches have been proposed for head and neck cancer segmentation. Some of them use standard convolutional neural network classifiers on patches with tailored pre- and postprocessing [27-31]. However, the U-Net

XSL•FO
RenderX

convolutional architecture [32] has shown promise in the area of deep learning–based medical image segmentation [33] and has also been applied to head and neck radiotherapy segmentation [34-47].

Despite the promise that deep learning offers, barriers remain in the application of auto-segmentation in radiotherapy planning. These include the absence of consensus on how *expert* performance is defined, the lack of available methods by which such human performance can be compared with that delivered by automated segmentation processes, and thus how the clinical acceptability of automated processes can be defined.

Objectives

In this paper, we address these challenges in defining comparison metrics and report a deep learning approach that delineates a wide range of important organs at risk in head and neck cancer radiotherapy scans. We aim to achieve this using a study design that includes (1) the introduction of a clinically meaningful performance metric for segmentation in radiotherapy planning, (2) a representative set of images acquired during routine clinical practice, (3) an unambiguous segmentation protocol for all organs, and (4) a segmentation of each test set image according to these protocols by 2 independent experts. In addition to the model's generalizability, as demonstrated on two distinct open-source data sets, by achieving performance equal to that of human experts on previously unseen patients from the same hospital site used for training, we aim to demonstrate the clinical applicability of our approach.

Methods

Data Sets

University College London Hospitals (UCLH) National Health Service (NHS) Foundation Trust serves an urban, mixed socioeconomic and ethnic population in central London, United Kingdom, and houses a specialist center for cancer treatment. Data were selected from a retrospective cohort of all-adult (aged >18 years) UCLH patients who underwent computed tomography (CT) scans to plan radical radiotherapy treatment for head and neck cancer between January 1, 2008, and March 20, 2016. Both initial CT images and rescans were included in the training data set. Patients with all tumor types, stages, and histological grades were considered for inclusion, as long as their CT scans were available in digital form and were of sufficient diagnostic quality. The standard CT pixel spacing was 0.976×0.976×2.5 mm, and scans with nonstandard spacing (with the exception of 1.25-mm spacing scans that were subsampled) were excluded to ensure consistent performance metrics during training. It should be noted that for the Cancer Imaging Archive (TCIA) test set, the in-plane pixel spacing was not used as an exclusion criterion, *i* ranged from 0.94 to 1.27 mm. For the public domain database for computational anatomy (PDDCA) test set, we included all scans, and the voxels varied between 2 to 3 mm in height and 0.98 to 1.27 mm in axial dimension. Patients' requests to not have their data shared for research were respected.

Of the 513 patients who underwent radiotherapy at UCLH within the given study dates, a total of 486 patients (94.7%; 838 scans; mean age 57 years; male 337, female 146, and gender unknown 3) met the inclusion criteria. Of note, no scans were excluded because of poor diagnostic quality. Scans from UCLH were split into a training set (389 patients; 663 scans), validation set (51 patients; 100 scans), and test set (46 patients; 75 scans). From the selected test set, 19 patients (21 scans) underwent adjudicated contouring described below. No patient was included in multiple data sets; in cases where multiple scans were present for a single patient, all were included in the same subset. Multiple scans present for a single patient reflect CT scans taken for the purpose of replanning radiotherapy owing to anatomical changes during the course of treatment. It is important for models to perform well in both scenarios as treatment naive and postradiotherapy organ at risk anatomies can differ. However, to avoid potential correlation between the same organs segmented twice in the same data set, care was taken to avoid this in the TCIA test set (described later in this section).

In total, 21 organs at risk were selected throughout the head and neck area to represent a wide range of anatomical regions. We used a combination of segmentations sourced from those used clinically at UCLH and additional segmentations performed in-house by trained radiographers.

We divided our UCLH data set into the following categories: (1) *training set*, used to train the model, a combination of UCLH clinical segmentations and in-house segmentations, some of which were only 2D slices (owing to the time required to segment larger organs manually, we initially relied heavily on sparse segmentations to make efficient use of the radiographers' time). (2) *UCLH validation set*: used to evaluate model performance and steer additional data set priorities, which used in-house segmentations only, as we did not want to overfit any clinical bias. (3) *UCLH test set*: our primary result set; each scan has every organ at risk labeled and was independently segmented from scratch by 2 radiographers before one of the pairs of scans (chosen arbitrarily) was reviewed and corrected by an experienced radiation oncologist.

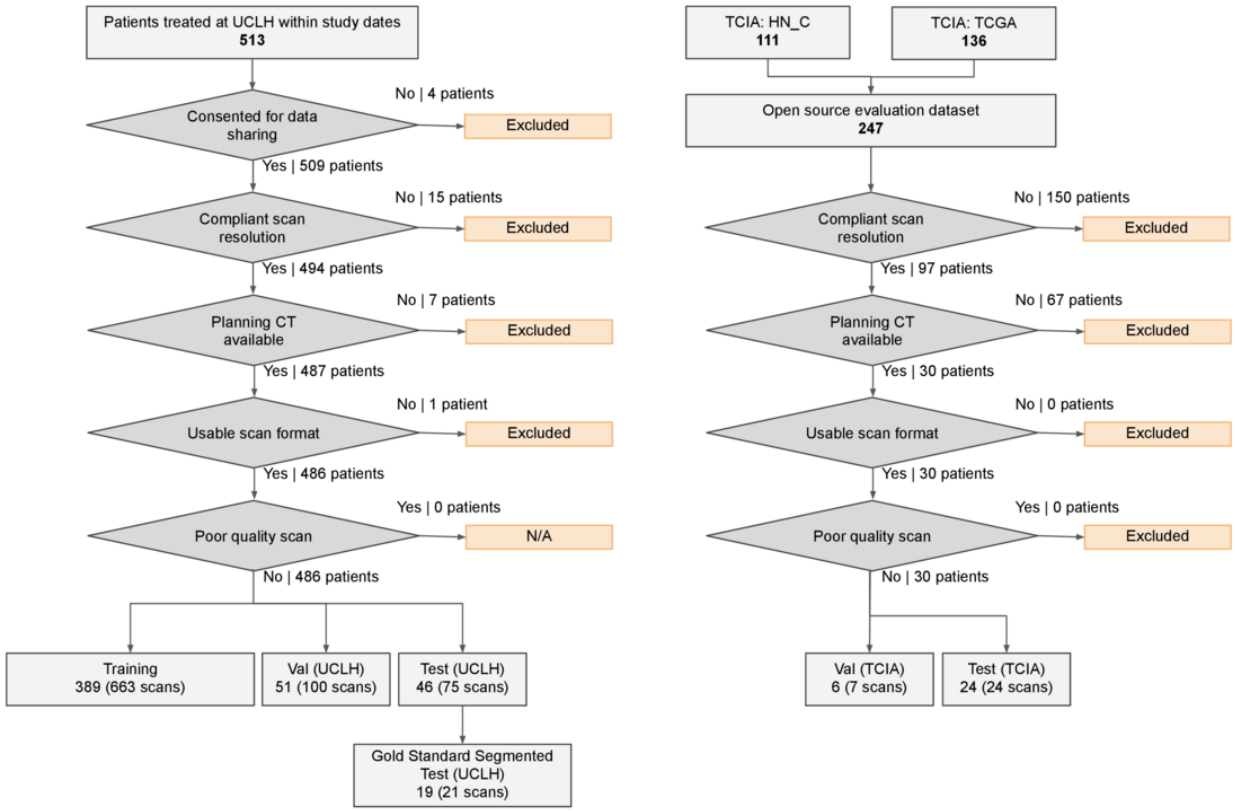
As these scans were taken from UCLH patients not present elsewhere, and to consider generalizability, we curated additional open-source CT scans available from The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma (TCGA-HNSC) and Head-Neck Cetuximab [48-50]. The open-source (category 4) TCIA validation set and (category 5) TCIA test set were both labeled in the same way as our UCLH test set.

Non-CT planning scans and those that did not meet the same slice thickness as the UCLH scans (2.5 mm) were excluded. These were then manually segmented in-house according to the Brouwer Atlas (the segmentation procedure is described in further detail in the *Clinical Labeling and Annotation* section [51]). We included 31 scans (22 Head-Neck Cetuximab and 9 TCGA-HNSC) that met these criteria, which we further split into validation (6 patients; 7 scans) and test (24 patients; 24 scans) sets (Figure 2). The original segmentations from the Head-Neck Cetuximab data set were not included; a consensus assessment by experienced radiographers and oncologists found the segmentations either nonconformant to the selected

XSL•FO
RenderX

segmentation protocol or below the quality that would be acceptable for clinical care. The original inclusion criteria for Head-Neck Cetuximab were patients with stage 3-4 carcinoma of the oropharynx, larynx, and hypopharynx, with a Zubrod performance of 0-1, and meeting predefined blood chemistry criteria between November 2005 and March 2009. The TCGA-HNSC data set included patients treated for head-neck squamous cell carcinoma, with no further restrictions being apparent [48,50].

Figure 2. Case selection from the University College London Hospitals and The Cancer Imaging Archive computed tomography data sets. A consort-style diagram demonstrating the application of inclusion and exclusion criteria to select the training, validation, and test sets used in this work. CT: computed tomography; HN_C: Head and Neck Carcinoma; N/A: not applicable; TCIA: The Cancer Imaging Archive; TCGA: The Cancer Genome Atlas; UCLH: University College London Hospitals; Val: validation.



All test sets were kept separate during model training and validation. Table 1 describes in detail the demographics and characteristics within the data sets; to obtain a balanced demographic in each of the tests, the validation and training data sets, we sampled randomly stratified splits and selected one that minimized the differences between the key demographics in each data set.

In addition, the (6) *PDDCA open-source data set* consisted of 15 patients selected from the Head-Neck Cetuximab open-source data set [48], owing to differences in selection criteria and test, validation, or training set allocation, five scans were present in both the TCIA and PDDCA test sets. This data set was used without further postprocessing and only accessed once to assess the volumetric Dice similarity coefficient (DSC) performance. The PDDCA test set differs from the TCIA test set in both the segmentation protocol and the axial slice thickness. The work by Raudaschl et al [25] provides more details on the data set characteristics and preprocessing.

Table 1 details the characteristics of these data sets and patient demographics.

XSL•FO
RenderX

Table 1. Data set characteristics[a].

Data set	UCLH[b]			TCIA[c]		PDDCA[d]
	Train	Validation	Test	Validation	Test	Test
Total scans (patients), n	663 (389)	100 (51)	21 (19)	7 (6)	24 (24)	15 (15)
Average patient age (years)	57.1	57.5	59.6	56.5	59.9	58.6
Sex, number of scans (number of patients)						
Female	207 (115)	36 (19)	7 (6)	2 (2)	2 (2)	2 (2)
Male	450 (271)	64 (32)	14 (13)	5 (4)	20 (20)	9 (9)
Unknown	6 (3)	0 (0)	0 (0)	0 (0)	2 (2)	4 (4)
Tumor site, number of scans (number of patients)						
Oropharynx	145 (86)	27 (15)	7 (6)	0 (0)	8 (8)	2 (2)
Lip, oral cavity, and pharynx	80 (52)	20 (8)	4 (4)	1 (1)	3 (3)	0 (0)
Tongue	53 (26)	8 (5)	1 (1)	2 (2)	7 (7)	0 (0)
Larynx	46 (31)	8 (3)	2 (2)	2 (2)	4 (4)	0 (0)
Nasopharynx	48 (24)	5 (3)	0 (0)	0 (0)	0 (0)	0 (0)
Head, face, and neck	37 (23)	8 (3)	1 (1)	0 (0)	0 (0)	0 (0)
Nasal cavity	32 (19)	2 (1)	1 (1)	0 (0)	0 (0)	0 (0)
Connective and soft tissue	37 (18)	2 (1)	1 (1)	0 (0)	0 (0)	0 (0)
Hypopharynx	17 (10)	1 (1)	0 (0)	2 (1)	1 (1)	0 (0)
Accessory sinus	10 (7)	2 (1)	0 (0)	0 (0)	0 (0)	0 (0)
Esophagus	6 (2)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
Other	33 (20)	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)
Unknown	119 (71)	16 (9)	4 (3)	0 (0)	0 (0)	13 (13)
Source, number of scans (number of patients)						
TCGA[e]	—[f]	—	—	2 (2)	7 (7)	0 (0)
HN_Cetux[g]	—	—	—	5 (4)	17 (17)	15 (15)
Site, number of scans (number of patients)						
UCLH	663 (389)	100 (51)	21 (19)	0 (0)	0 (0)	0 (0)
MD Anderson Cancer Center	0 (0)	0 (0)	0 (0)	2 (2)	7 (7)	0 (0)
Unknown (US)	0 (0)	0 (0)	0 (0)	5 (4)	17 (17)	15 (15)

[a]Tumor sites were derived from International Classification of Diseases codes. The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma [52] is an open-source data set hosted on The Cancer Imaging Archive (TCIA). Head-Neck Cetuximab is an open-source data set hosted on TCIA [53]. Public Domain Database for Computational Anatomy data set released as part of the 2015 challenge in the segmentation of head and neck anatomy at the International Conference on Medical Image Computing and Computer Assisted Intervention.

[b]UCLH: University College London Hospitals.

[c]TCIA: The Cancer Imaging Archive.

[d]PDDCA: Public Domain Database for Computational Anatomy.

[e]TCGA: The Cancer Genome Atlas Program.

[f]The University College London Hospitals (UCLH) data set was sourced entirely from UCLH.

[g]HN_Cetux: Head-Neck Cetuximab.

Clinical Taxonomy

To select the organs at risk to be included in the study, we used the Brouwer Atlas (consensus guidelines for delineating organs at risk for head and neck radiotherapy, defined by an international panel of radiation oncologists [51]). From this, we excluded those regions that required additional magnetic resonance imaging for segmentation, those that were not relevant to routine head and neck radiotherapy, or those that were not used clinically at UCLH. This resulted in a set of 21 organs at risk (Table 2).

XSL•FO
RenderX

Table 2. Taxonomy of segmentation regions.

Organ at risk	Total number of labeled slices included	Anatomical landmarks and definition
Brain	11,476	Sits inside the cranium and includes all brain vessels excluding the brainstem and optic chiasm.
Brainstem	34,794	The posterior aspect of the brain including the midbrain, pons, and medulla oblongata. Extending inferior from the lateral ventricles to the tip of the dens at C2. It is structurally continuous with the spinal cord.
Cochlea-left	4526	Embedded in the temporal bone and lateral to the internal auditory meatus.
Cochlea-right	4754	Embedded in the temporal bone and lateral to the internal auditory meatus.
Lacrimal-left	17,186	Concave-shaped gland located at the superolateral aspect of the orbit.
Lacrimal-right	17,788	Concave-shaped gland located at the superolateral aspect of the orbit.
Lens-left	3006	An oval structure that sits within the anterior segment of the orbit. Can be variable in position but never sitting posterior beyond the level of the outer canthus.
Lens-right	3354	An oval structure that sits within the anterior segment of the orbit. Can be variable in position but never sitting posterior beyond the level of the outer canthus.
Lung-left	8340	Encompassed by the thoracic cavity adjacent to the lateral aspect of the mediastinum, extending from the first rib to the diaphragm excluding the carina.
Lung-right	9158	Encompassed by the thoracic cavity adjacent to the lateral aspect of the mediastinum, extending from the first rib to the diaphragm excluding the carina.
Mandible	25,074	The entire mandible bone including the temporomandibular joint, ramus, and body, excluding the teeth. The mandible joins to the inferior aspect of the temporal bone and forms the entire lower jaw.
Optic-nerve-left	3458	A 2 to 5 mm thick nerve that runs from the posterior aspect of the eye, through the optic canal and ends at the lateral aspect of the optic chiasm.
Optic-nerve-right	3012	A 2 to 5 mm thick nerve that runs from the posterior aspect of the eye, through the optic canal and ends at the lateral aspect of the optic chiasm.
Orbit-left	8538	Spherical organ sitting within the orbital cavity. Includes the vitreous humor, retina, cornea, and lens with the optic nerve attached posteriorly.
Orbit-right	8242	Spherical organ sitting within the orbital cavity. Includes the vitreous humor, retina, cornea, and lens with the optic nerve attached posteriorly.
Parotid-left	8984	Multi-lobed salivary gland wrapped around the mandibular ramus. Extends medially to the styloid process and parapharyngeal space. Laterally extending to the subcutaneous fat. Posteriorly extending to the sternocleidomastoid muscle. Anterior extending to posterior border of the mandible bone and masseter muscle. In cases where the retromandibular vein is encapsulated by parotid, this is included in the segmentation.
Parotid-right	11,752	Multi-lobed salivary gland wrapped around the mandibular ramus. Extends medially to the styloid process and parapharyngeal space. Laterally extending to the subcutaneous fat. Posteriorly extending to the sternocleidomastoid muscle. Anterior extending to posterior border of the mandible bone and masseter muscle. In cases where the retromandibular vein is encapsulated by parotid this is included in the segmentation.
Spinal-canal	37,000	Hollow cavity that runs through the foramen of the vertebrae, extending from the base of skull to the end of the sacrum.
Spinal-cord	37,096	Sits inside the spinal canal and extends from the level of the foramen magnum to the bottom of L2.
Submandibular-left	10,652	Sits within the submandibular portion of the anterior triangle of the neck, making up the floor of the mouth and extending both superior and inferior to the posterior aspect of the mandible and is limited laterally by the mandible and medially by the hypoglossal muscle.
Submandibular-right	10,716	Sits within the submandibular portion of the anterior triangle of the neck, making up the floor of the mouth and extending both superior and inferior to the posterior aspect of the mandible and is limited laterally by the mandible and medially by the hypoglossal muscle.

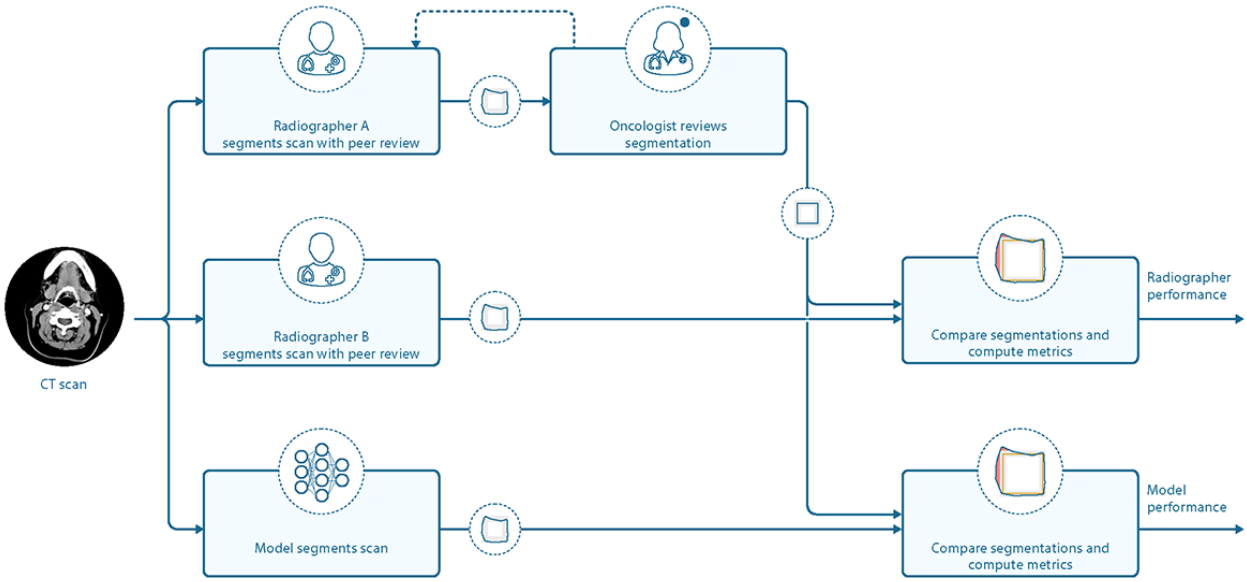
XSL•FO
RenderX

Clinical Labeling and Annotation

Owing to the large variability of segmentation protocols used and annotation quality in the UCLH data set, all segmentations from all scans selected for inclusion in the training set were manually reviewed by a radiographer with at least 4 years of experience in the segmentation of head and neck organs at risk. Volumes that did not conform to the Brouwer Atlas were excluded from the training. To increase the number of training examples, additional axial slices were randomly selected for further manual organ at risk segmentations to be added based on model performance or perceived imbalances in the data set. These were then produced by a radiographer with at least 4 years of experience in head and neck radiotherapy, arbitrated by a second radiographer with the same level of experience. The total number of examples from the original UCLH segmentations and additional slices are provided in Table 2.

For the TCIA test and validation sets, the original dense segmentations were not used owing to poor adherence to the chosen study protocol. To produce the ground truth labels, the full volumes of all 21 organs at risk included in the study were segmented. This was done initially by a radiographer with at least 4 years of experience in the segmentation of head and neck organs at risk and then arbitrated by a second radiographer with similar experience. Further arbitration was then performed by a radiation oncologist with at least 5 years of postcertification experience in head and neck radiotherapy. The same process was repeated with 2 additional radiographers working independently, but after peer arbitration, these segmentations were not reviewed by an oncologist; rather, they became the human reference to which the model was compared. This is schematically shown in Figure 3. Before participation, all radiographers and oncologists were required to study the Brouwer Atlas for head and neck organ at risk segmentation [51] and demonstrate competence in adhering to these guidelines.

Figure 3. Process for the segmentation of ground truth and radiographer organs at risk volumes. The flowchart illustrates how the ground truth segmentations were created and compared with independent radiographer segmentations and the model. For the ground truth, each computed tomography scan in The Cancer Imaging Archive test set was segmented first by a radiographer and peer reviewed by a second radiographer. This then went through one or more iterations of review and editing with a specialist oncologist before creating a ground truth used to compare with the segmentations produced by both the model and additional radiographer. CT: computed tomography.

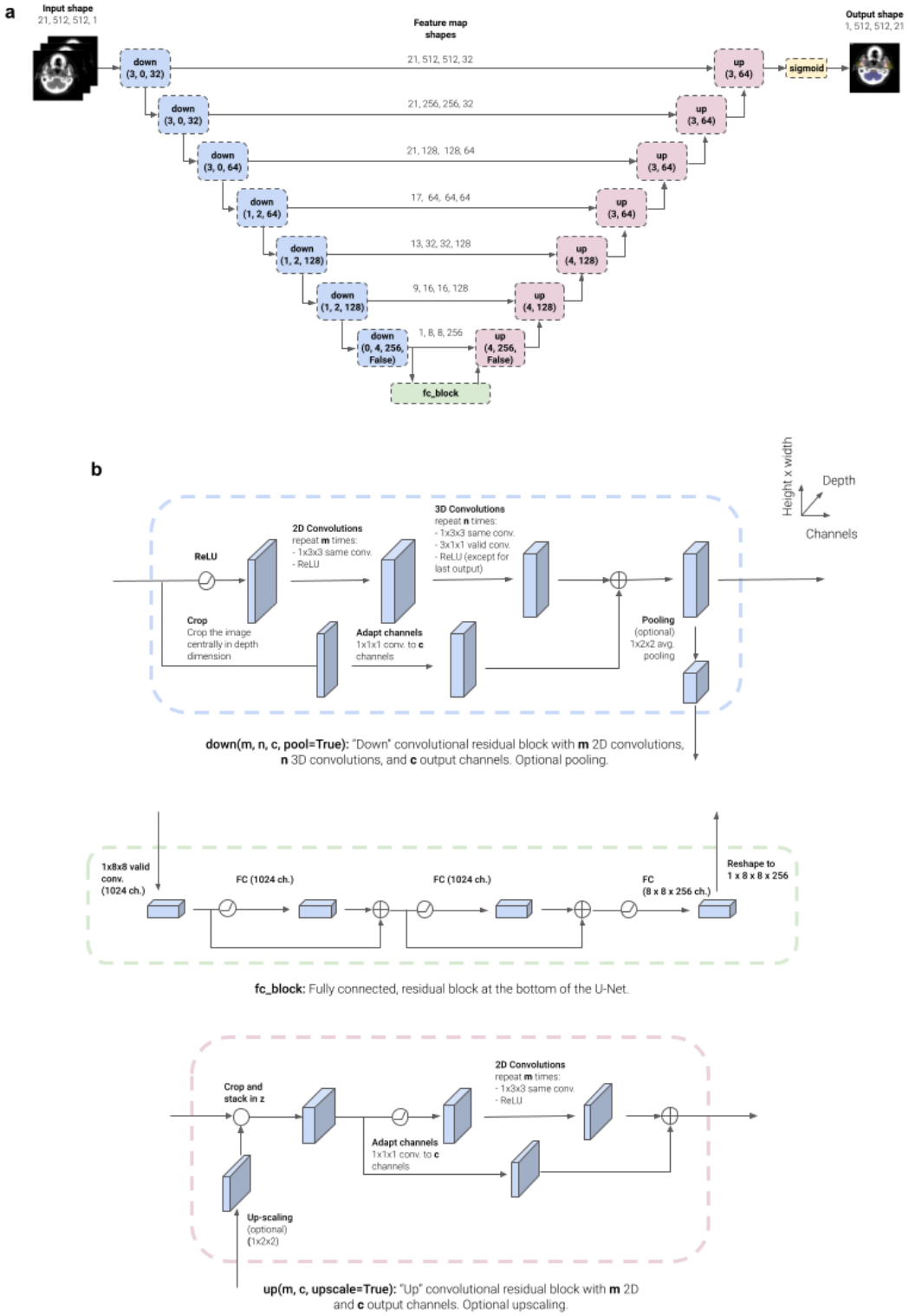


Model Architecture

We used a residual 3D U-Net architecture with 8 levels (Figure 4). Our network takes in a CT volume (single channel) and outputs a segmentation mask with 21 channels, where each channel contains a binary segmentation mask for a specific organ at risk. The network consists of 7 residual convolutional blocks in the downward path, a residual fully connected block at the bottom, and 7 residual convolutional blocks in the upward path. A 1×1×1 convolution layer with sigmoidal activation produces the final output in the original resolution of the input image. Each predicted slice had 21 slices of context. The 21-slice context (ie, 21 × 2.5 mm=52.5 mm) was found to provide the optimal context. This is not the case with the 21 organs at risk used in this study.

XSL•FO
RenderX

Figure 4. 3D U-Net model architecture. (a) At training time, the model receives 21 contiguous computed tomography slices, which are processed through a series of "down" blocks, a fully connected block, and a series of "up" blocks to create a segmentation prediction. (b) A detailed view of the convolutional residual down and up blocks and the residual fully connected block.



We trained our network with a regularized top-*k*-percent, pixel-wise, binary, cross-entropy loss [54]; for each output channel, the top-*k* loss selects only the *k*% most difficult pixels (those with the highest binary cross-entropy) and only adds their contribution to the total loss. This speeds up training and helps the network to tackle the large class imbalance and to focus on difficult examples.

We regularized the model using standard L2 weight regularization with scale 10^{-6} and extensive data augmentation using random in-plane (ie, in *x* and *y* directions only) translation, rotation, scaling, shearing, mirroring, elastic deformations, and pixel-wise noise. We used uniform translations between −32 and 32 pixels, uniform rotations between −9° and 9°, uniform scaling factors between 0.8° and 1.2°, and uniform shear factors

XSL•FO
RenderX

between −0.1 and 0.1. We mirrored the images (and adjusted the corresponding left and right labels) with a probability of 0.5. We performed elastic deformations by placing random displacement vectors (SD 5 mm, in-plane displacements only) on a control point grid with 100×100×100 mm spacing and by deriving the dense deformation field using cubic b-spline interpolation. In the implementation, all spatial transformations are first combined to a dense deformation field, which is then applied to the image using bilinear interpolation and extrapolation with zero padding. We added zero-mean Gaussian intensity noise independently to each pixel with an SD of 20 Hounsfield units.

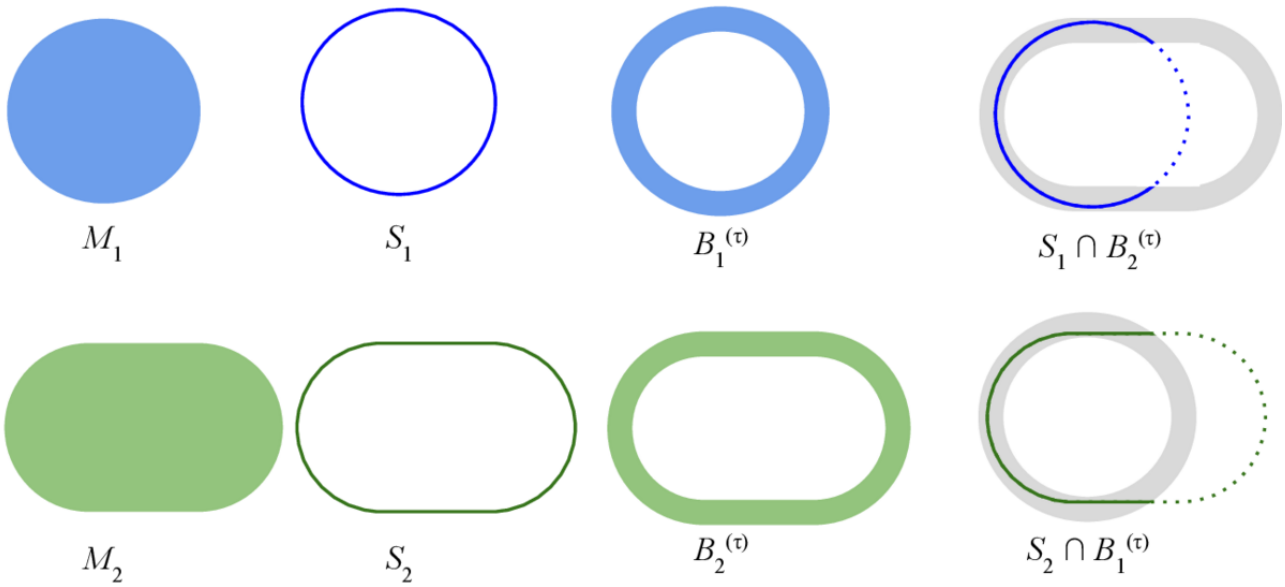
We trained the model with the Adam optimizer [53] for 120,000 steps and a batch size of 32 (32 graphical processing units) using synchronous stochastic gradient descent. We used an initial learning rate of 10^{-4} and scaled the learning rate by 1/2, 1/8, 1/64, and 1/256 at time steps of 24,000, 60,000, 108,000, and 114,000, respectively.

We used the validation set to select the model that performed at over 95% for most organs at risk according to our chosen surface DSC performance metric, breaking ties by preferring better performance on more clinically impactful organs at risk and the absolute performance obtained.

Performance Metrics

All performance metrics are reported for each organ independently (eg, separately for just the left parotid), so we only need to deal with binary masks (eg, a left parotid voxel and a non–left-parotid voxel). Masks are defined as a subset of , that is, (Figure 5).

Figure 5. Illustrations of masks, surfaces, border regions, and the "overlapping" surface at tolerance τ.



The volume of a mask is denoted as , with



With this notation, the standard (volumetric) DSC for two given masks M_1 and M_2 and can be written as:



In the case of sparse ground truth segmentations (ie, only a few slices of the CT scan are labeled), we estimate the volumetric DSC by aggregating data from labeled voxels across multiple scans and patients as



where the mask $M_{1,p}$ and the labeled region L_p represent the sparse ground truth segmentation for a patient p and the mask $M_{2,p}$ is the full volume predicted segmentation for the patient p.

Owing to the shortcomings of the volumetric DSC metric for the presented radiotherapy use case, we introduced the *surface DSC* metric, which assesses the overlap of two surfaces (at a specified tolerance) instead of the overlap of two volumes (see *Results* section). A surface is the border of a mask, , and the area of the surface is denoted as



where is a point on the surface using arbitrary parameterization. The mapping from this parameterization to a point in is denoted as , that is, . With this we can define the border region , for the surface S_i, at a given tolerance τ as (Figure 5)



Using these definitions, we can write the *surface DSC at tolerance* τ as



using an informal notation for the intersection of the surface with the boundary, that is,

XSL•FO
RenderX

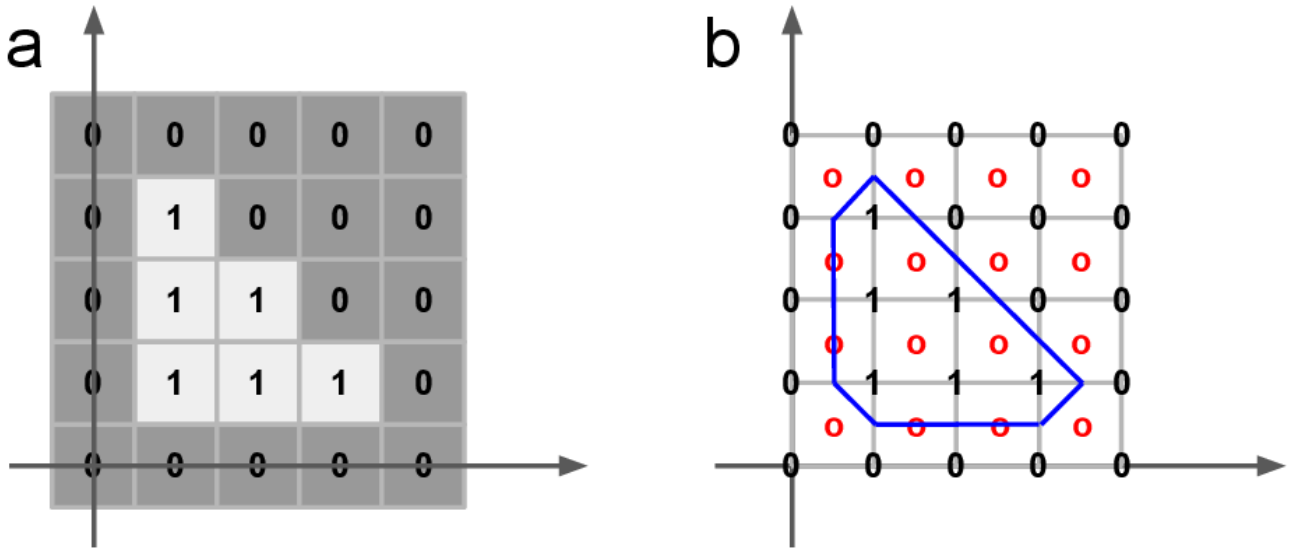
Implementation of Surface DSC

The computation of surface integrals on sampled images is not straightforward, especially for medical images, where the voxel spacing is usually not equal in all 3 dimensions. The common approximation of the integral by counting the surface voxels can lead to substantial systematic errors.

Another common challenge is the representation of a surface with voxels. As the surface of a binary mask is located between voxels, a definition of *surface voxels* in the raster-space of the image introduces a bias: using foreground voxels to represent the surface leads to an underestimation of the surface, whereas the use of background voxels leads to an overestimation.

Our proposed implementation uses a surface representation that provides less-biased estimates but still allows us to compute the performance metrics with linear complexity O(N), with N: number of voxels). We placed the surface points between the voxels on a raster that is shifted by half of the raster spacing on each axis (see Figure 6 for a 2D illustration).

Figure 6. 2D illustration of the implementation of the surface Dice similarity coefficient. (a) A binary mask displayed as an image. The origin of the image raster is (0,0). (b) The surface points (red circles) are located in a raster that is shifted half of the raster spacing on each axis. Each surface point has 4 neighbors in 2D (8 neighbors in 3D). The local contour (blue line) assigned to each surface point (red circle) depends on the neighbor constellation.



For 3D images, each point in the raster has 8 neighboring voxels. As we analyzed binary masks, there are only 2^8=256 possible neighbor constellations. For each of these constellations, we computed the resulting triangles using the marching cube triangulation [55,56] and stored the surface area of the triangles (in mm^2) in a look-up table. With this look-up table, we then created a surface image (on the above-mentioned raster) that contains zeros at positions that have 8 identical neighbors or the local surface area at all positions that have both foreground and background neighbors. These surface images were created for masks M_1 and M_2. In addition, we created a distance map from each of these surface images using the distance transform algorithm [57]. Iterating over the nonzero elements in the first surface image and looking up the distance from the other surface in the corresponding distance map allows the creation of a list of tuples (surface element area and distance from other surfaces). From this list, we can easily compute the surface area by summing the area of the surface elements that are within the tolerance. To account for the quantized distances, there is only a discrete set of distances between voxels in a 3D raster with spacing (d_1, d_2, d_3)—we also rounded the tolerance to the nearest neighbor in set D for each image before computing the surface DSC. Our open-source implementation of surface DSC provides more details.

Results

Selecting Clinically Representative Data Sets

Data sets are described in detail in the Methods section. In brief, the first data set was a representative sample of CT scans used to plan curative-intent radiotherapy of head and neck cancer for patients at UCLH NHS Foundation Trust, a single high-volume center. We performed iterative cycles of model development using the UCLH scans (*training* and *validation* subsets), taking the performance on a previously unseen subset (*test*) as our primary outcome.

It is also important to demonstrate a model's generalizability to data from previously unseen demographics and distributions. To do this, we curated the test and validation data sets of open-source CT scans. These were collected from the *TCIA test set* [48-50] and the *PDDCA data set* released as part of the 2015 challenge (*PDDCA test set* [25]).

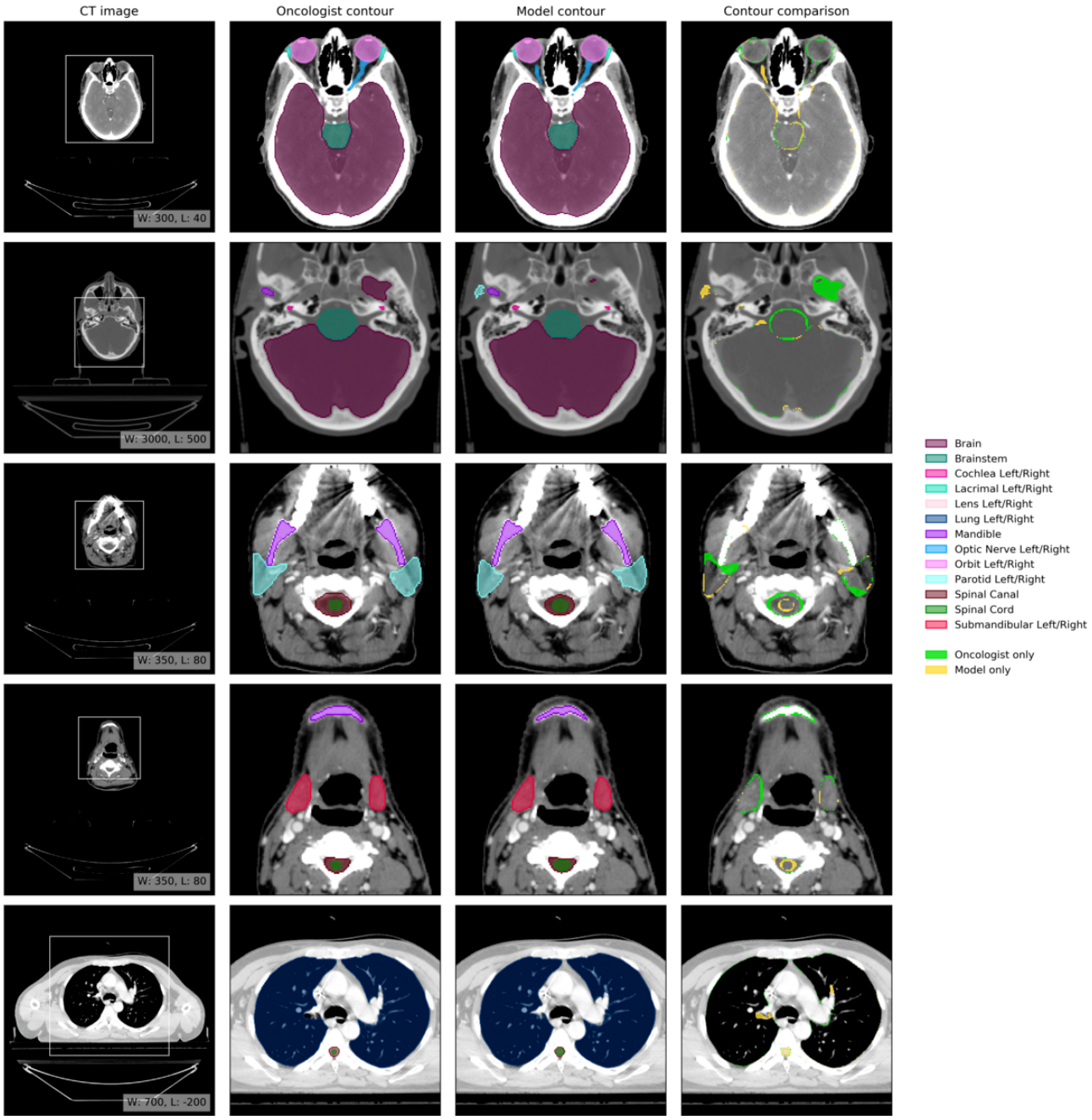
Table 1 details the characteristics of these data sets and their patient demographics. Ethnicity and protected-group status are not reported, as this information was not available in the source systems. In total, 21 organs at risk were selected to represent a wide range of anatomical regions throughout the head and neck. To provide a human clinical comparison for the algorithm, each case was manually segmented by a single radiographer with arbitration by a second radiographer. This was compared with

XSL•FO
RenderX

our study's *gold standard* ground truth graded by 2 other radiographers and arbitrated by one of 2 independent specialist oncologists, each with a minimum of 4 years specialist experience in radiotherapy treatment planning for patients with head and neck cancer.

An example of model performance is shown in Figure 7, two further randomly selected UCLH set scans are shown in Figures S1 and S2 of Multimedia Appendix 1 [19-31,34-46,56-90]. Three randomly selected TCIA set scans are shown in Figures S3, S4 and S5 of Multimedia Appendix 1 to visually demonstrate the model's generalizability. We compared our performance (model vs oncologist) to radiographer performance (radiographer vs oncologist). For more information on data set selection and inclusion and exclusion criteria for patients and organs at risk, see the *Methods* section.

Figure 7. Example results. Computed tomography (CT) image: axial slices at 5 representative levels from the raw CT scan of a male patient aged 55-59 years were selected from the University College London Hospitals data set (patient 20). These were selected to best demonstrate the organs at risks included in the work. The levels shown as 2D slices have been selected to demonstrate all 21 organs at risks included in this study. The window leveling has been adjusted for each to best display the anatomy present. Oncologist contour: the ground truth segmentation, as defined by experienced radiographers and arbitrated by a head and neck specialist oncologist. Model contour: segmentations produced by our model. Contour comparison: contoured by oncologist only (green region) or model only (yellow region). Best viewed on a display. CT: computed tomography.



A New Metric for Assessing Clinical Performance

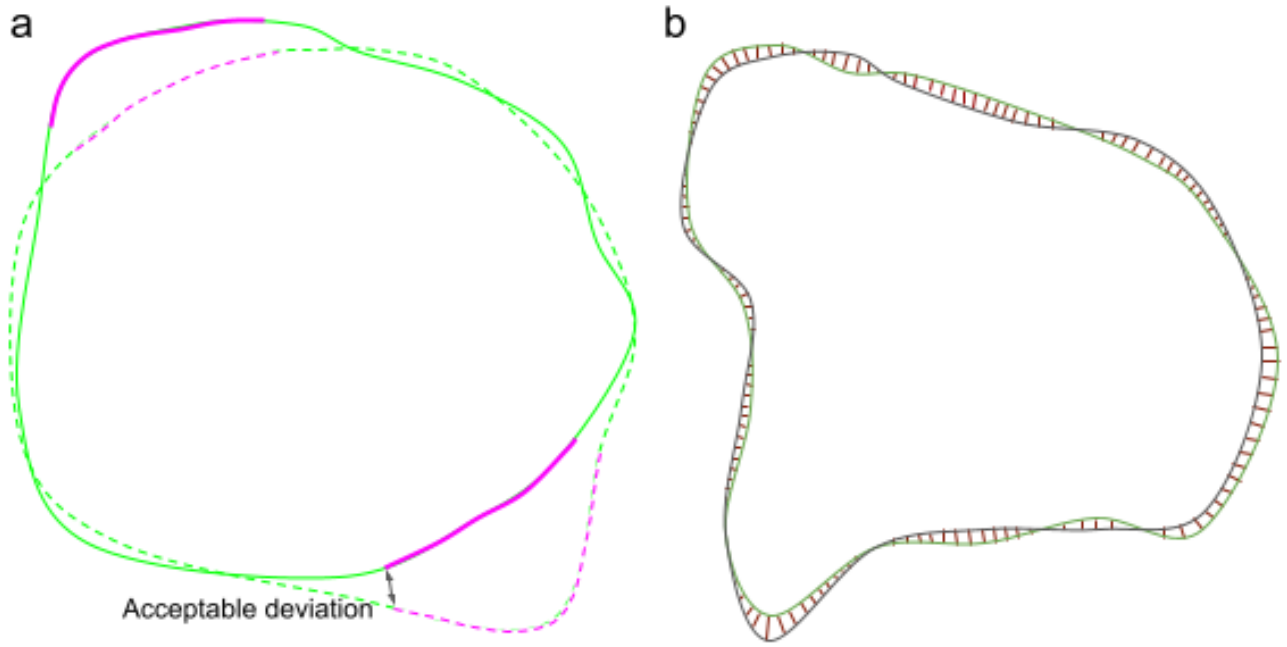
In routine clinical care, algorithm-derived segmentation is reviewed and potentially corrected by a human expert, just as those created by radiographers currently are. Segmentation performance is thus best assessed by determining the fraction of the surface that needs to be redrawn. The standard volumetric

XSL•FO
RenderX

DSC [91] is not well suited to this because it weighs all regions of misplaced delineation equally and independently of their distance from the surface. For example, two inaccurate segmentations could have a similar volumetric DSC score if one were to deviate from the correct surface boundary by a small amount in many places, whereas the other had a large deviation at a single point. Correcting the former would likely take a considerable amount of time as it would require redrawing almost all of the boundary, whereas the latter could be corrected much faster, potentially with a single edit action.

For quantitative analysis, we therefore introduced a new segmentation performance metric, the *surface DSC* (Figure 8), which assesses the overlap of two surfaces (at a specified tolerance) instead of the overlap of two volumes. This provides a measure of agreement between the surfaces of two structures, which is where most of the human effort in correcting is usually expended. In doing so, we also addressed the volumetric DSC's bias toward large organs at risk, where the large (and mostly trivial) internal volume accounts for a much larger proportion of the score.

Figure 8. Surface Dice similarity coefficient performance metric. (a) Illustration of the computation of the surface Dice similarity coefficient. Continuous line: predicted surface. Dashed line: ground truth surface. Black arrow: the maximum margin of deviation that may be tolerated without penalty, hereafter referred to by τ. Note that in our use case each organ at risk has an independently calculated value for τ. Green: acceptable surface parts (distance between surfaces ≤τ). Pink: unacceptable regions of the surfaces (distance between surfaces ≤τ). The proposed surface Dice similarity coefficient metric reports the good surface parts compared with the total surface (sum of predicted surface area and ground truth surface area). (b) Illustration of the determination of the organ-specific tolerance. Green: segmentation of an organ by oncologist A. Black: segmentation by oncologist B. Red: distances between the surfaces.



When evaluating the surface DSC, we must define a threshold within which the variation is clinically acceptable. To do this, we first defined the organ-specific tolerances (in mm) as a parameter of the proposed metric, τ. We computed these acceptable tolerances for each organ by measuring the interobserver variation in segmentations between 3 different consultant oncologists (each with over 10 years of experience in organ at risk delineation) on the validation subset of TCIA images.

To penalize both false-negative and false-positive parts of the predicted surface, our proposed metrics measure both the nonsymmetric distances between the surfaces and then normalize them by the combined surface area. Similar to volumetric DSC, the surface DSC ranges from 0 (no overlap) to 1 (perfect overlap).

This means that approximately 95% of the surface was properly outlined (ie, within τ mm of the correct boundary), whereas 5% needs to be corrected. There is no consensus as to what constitutes a nonsignificant variation in such a segmentation. Thus, we selected a surface DSC of 0.95, a stringency that likely far exceeds the expert oncologist intrarater concordance [19,92]. For a more formal definition and implementation, see the *Methods* section.

Model Performance

Model performance was evaluated alongside that of therapeutic radiographers (each with at least 4 years of experience) segmenting the test set of UCLH images independently of the oncologist-reviewed scans (which we used as our ground truth).

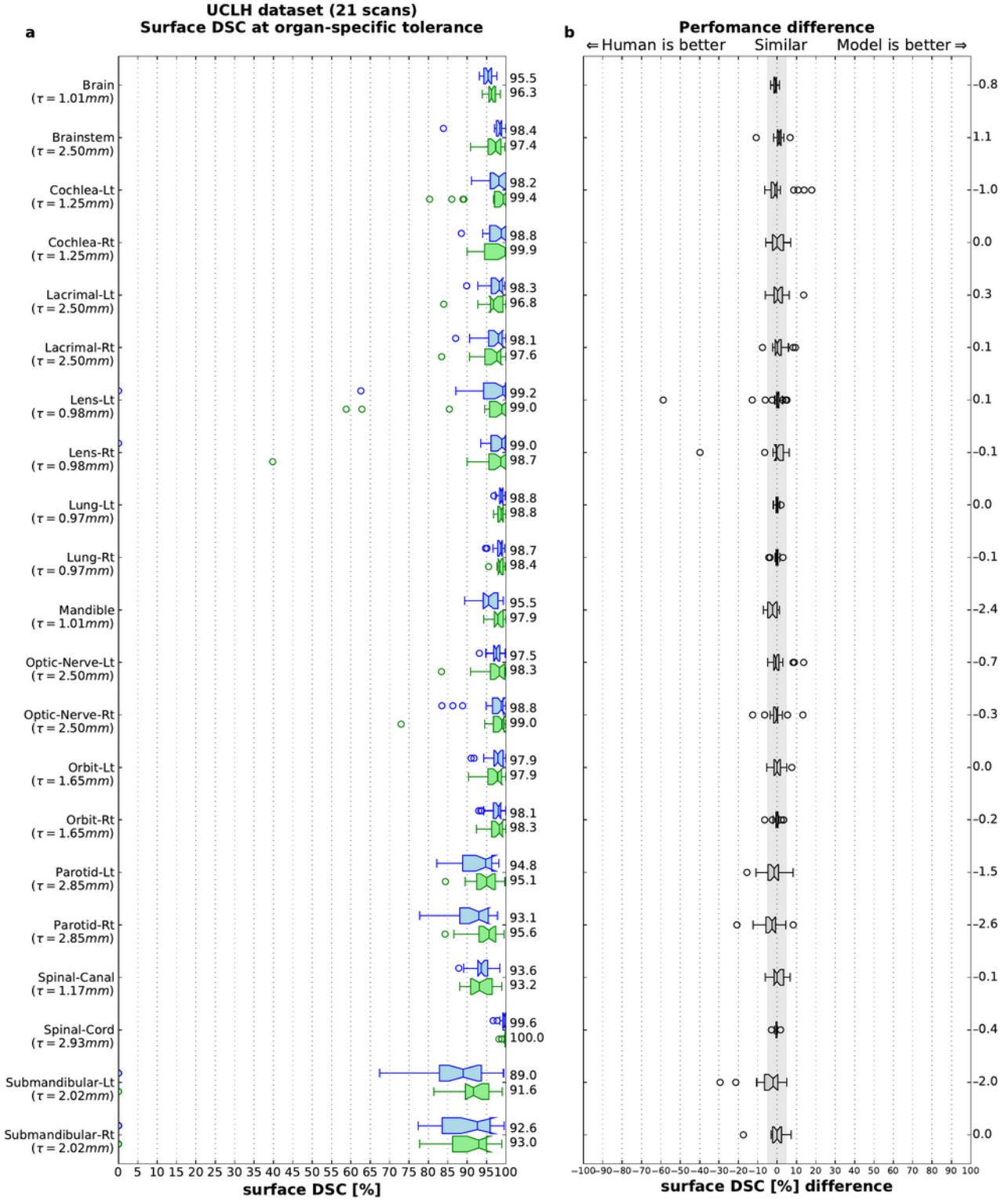
The model performed similarly to humans. For all organs at risk studied, there was no clinically meaningful difference between the deep learning model's segmentations and those of the radiographers (Figure 9 and Tables S1 and S2, Multimedia Appendix 1). For details on the number of labelled scans in the UCLH test set, see Table S3 in Multimedia Appendix 1.

To investigate the generalizability of our model, we additionally evaluated the performance of open-source scans (*TCIA test set*). These were collected from sites in the United States, where

XSL•FO
RenderX

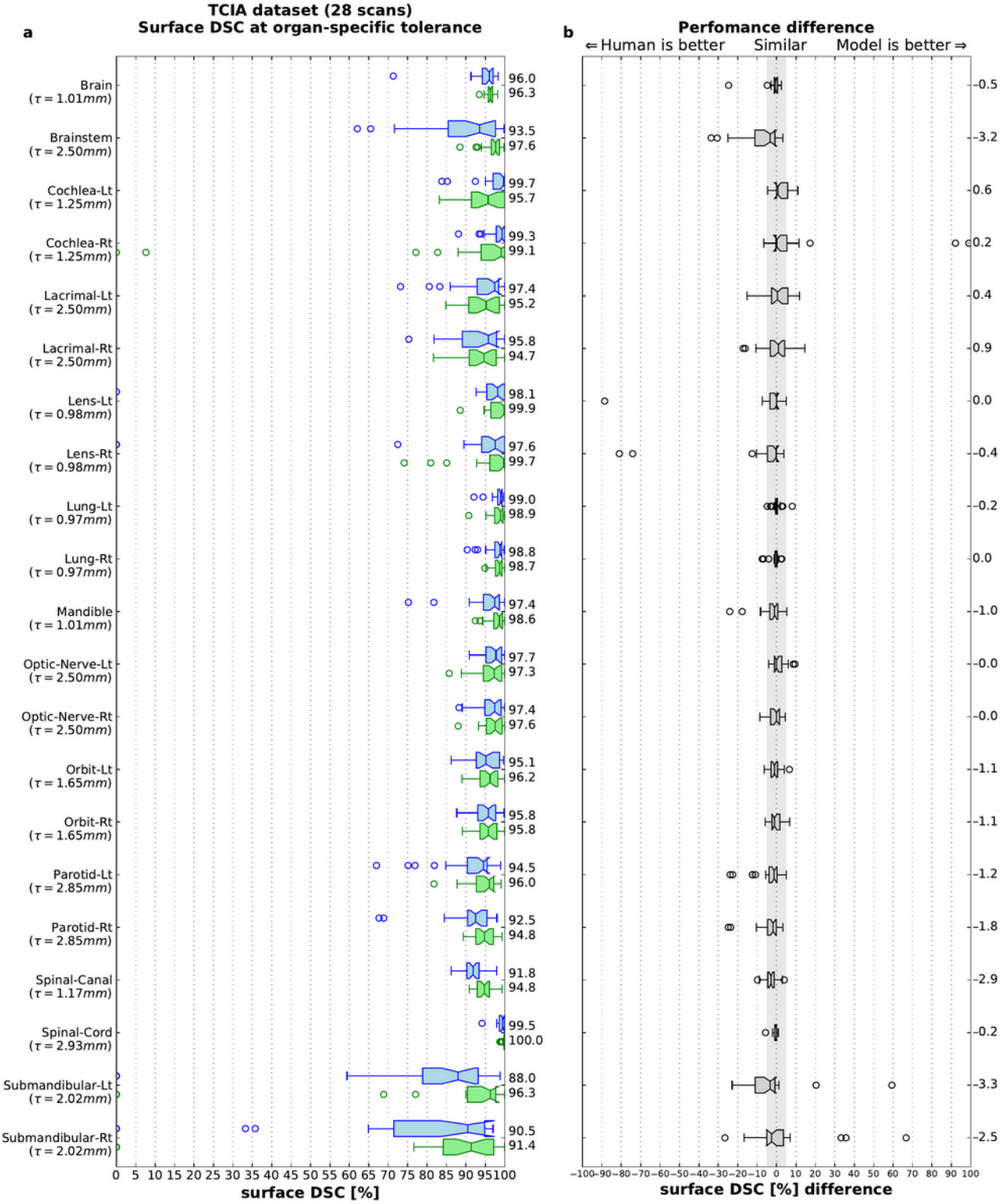
patient demographics, clinical pathways for radiotherapy, and scanner type and parameters differed from our UK training set in meaningful ways. Nevertheless, model performance was preserved, and in 90% (19/21) organs at risk, the model was performed within the threshold defined for human variability (Figure 10). The fact that performance in 2 organs at risk (brainstem and right lens) was less than that in UK data may relate to issues of image quality in several TCIA test set scans.

Figure 9. University College London Hospitals (UCLH) test set: quantitative performance of the model in comparison with radiographers. (a) The model achieves a surface Dice similarity coefficient similar to humans in all 21 organs at risk (on the UCLH held out test set) when compared with the gold standard for each organ at an organ-specific tolerance τ. Blue: our model; green: radiographers. (b) Performance difference between the model and the radiographers. Each blue dot represents a model-radiographer pair. The gray area highlights nonsubstantial differences (−5% to +5%). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers indicate most extreme, nonoutlier data points. Where data lie outside, an IQR of 1.5 is represented as a circular flier. The notches represent the 95% CI around the median. DSC: Dice similarity coefficient; UCLH: University College London Hospitals.



XSL•FO
RenderX

Figure 10. Model generalizability to an independent test set from The Cancer Imaging Archive (TCIA). Quantitative performance of the model on TCIA test set in comparison with radiographers. (a) Surface Dice similarity coefficient (on the TCIA open-source test set) for the segmentations compared with the gold standard for each organ at an organ-specific tolerance τ. Blue: our model, green: radiographers. (b) Performance difference between the model and the radiographers. Each blue dot represents a model-radiographer pair. Red lines show the mean difference. The gray area highlights nonsubstantial differences (−5% to +5%). The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers indicate most extreme, nonoutlier data points. Where data lie outside, an IQR of 1.5 is represented as a circular flier. The notches represent the 95% CI around the median. DSC: Dice similarity coefficient; TCIA: The Cancer Imaging Archive.



For more detailed results demonstrating surface DSC and volumetric DSC for each individual patient from the TCIA test set, see Table S4 and Table S5, respectively, in Multimedia Appendix 1.

Comparison With Previous Work

An accurate quantitative comparison with previously published literature is difficult because of inherent differences in definitions of ground truth segmentations and varied processes

XSL•FO
RenderX

of arbitration and consensus building. Given that the use of surface DSC is novel in this study, we also reported the standard volumetric DSC scores achieved by our algorithm (despite the shortcomings of this method) so that our results can be directly compared with those in the existing literature. An overview of past papers that have reported mean volumetric DSC for unedited automatic delineation of head and neck CT organs at risk can be found in Table S6, Multimedia Appendix 1. Each used different data sets, scanning parameters, and labeling protocols, meaning that the resulting volumetric DSC results varied significantly. No study, other than ours, segmented the lacrimal glands. We compared these results with those obtained when we applied our model to three different data sets: the TCIA open-source test set, an additional test set from the original UCLH data set (*UCLH test set*) and the data set released by the PDDCA as part of the 2015 Medical Image Computing and Computer Assisted Intervention head and neck radiotherapy organ at risk segmentation challenge (*PDDCA test set* [25]). To contextualize the performance of our model, radiographer performance is shown on the TCIA test set, and oncologist interobserver variation is shown on the UCLH test set.

Although not the primary test set, we nevertheless present per-patient surface DSC and volumetric DSC for the PDDCA test set in Table S7 and Table S8 in Multimedia Appendix 1, respectively.

Discussion

Principal Findings

We demonstrated an automated deep learning–based segmentation algorithm that can perform as well as experienced radiographers for head and neck radiotherapy planning. Our model was developed using CT scans derived from routine clinical practice and therefore should be applicable in a hospital setting for the segmentation of organs at risk, routine radiation therapy quality assurance peer review, and in reducing the associated variability between different specialists and radiotherapy centers [93].

Clinical applicability must be supported not only by high model performance but also by evidence of model generalizability to new external data sets. To achieve this, we presented these results on three separate test sets, one of which (the PDDCA test set) uses a different segmentation protocol. In this study, performance in most organs at risk was maintained when tested on scans taken from a range of previously unseen international sites. Although these scans varied in patient demographics, scanning protocol, device manufacturer, and image quality, the model still achieved human performance on 19 of the 21 organs at risk studied; only the right lens and brainstem were below radiographer performance. For these organs at risk, the performance of the model might have been lower than expert performance owing to lower image quality. This is particularly evident for the right lens, where the anatomical borders were quite indistinct in some TCIA test set cases, thus preventing full segmentation by the model (Figure S6, Multimedia Appendix 1). Moreover, a precise CT definition of the brainstem's proximal and distal boundaries is lacking, a factor that might have contributed to labeling variability and thus to decreased model performance. Finally, demographic bias may have resulted from the TCIA data set selection for cases of more advanced head and neck cancer [48] or from variability in the training data [10].

One major contribution of this paper is the presentation of a performance measure that represents the clinical task of organ at risk correction. In the first preprint of this work, we introduced surface DSC [70], a metric conceived to be sensitive to clinically significant errors in organ at risk delineation. Surface DSC has recently been shown to be more strongly correlated with the amount of time required to correct segmentation for clinical use than traditional metrics, including volumetric DSC [94,95]. Small deviations in organ at risk border placement can have a potentially serious impact, increasing the risk of debilitating side effects for the patient. Misplacement by only a small offset may thus require the entire region to be redrawn, and in such cases, an automated segmentation algorithm may offer no time savings. Volumetric DSC is relatively insensitive to such small changes in large organs, as the absolute overlap is also large. Difficulties identifying the exact borders of smaller organs can result in large differences in volumetric DSC, even if these differences are not clinically relevant in terms of their effect on radiotherapy treatment. By strongly penalizing border placement outside a tolerance determined by consultant oncologists, the surface DSC metric resolves these issues.

Although volumetric DSC is therefore not representative of clinical consequences, it remains to be the most popular metric for evaluating segmentation models and therefore the only metric that allows comparison with previously published works. In recent years, fully convolutional networks have become the most popular and successful methodology for organ at risk segmentation in head and neck CT for de novo radiotherapy planning [40-45,58-69]. Although not directly comparable owing to different data sets and labeling protocols, our volumetric DSC results compare favorably with the existing published literature for many of the organs at risk (see Table S6 and Figure S7, Multimedia Appendix 1, for more details on this and other prior publications). In organs at risk with inferior volumetric DSC scores compared with the published literature, both our model and human radiographers achieved similar scores. This suggests that current and previously published results are difficult to compare, either because of the inclusion of more difficult cases than previous studies or because of different segmentation and scanning protocols. To allow more objective comparisons of different segmentation methods, we made our labeled TCIA data sets freely available to the academic community (see the Acknowledgments section on data availability). At least 11 auto-segmentation software solutions are currently available commercially, with varying claims regarding their potential to lower segmentation time during radiotherapy planning [96]. The principal factor that determines whether automatic segmentation is time saving during the radiotherapy workflow is the degree to which automated segmentations require correction by oncologists.

The wide variability in state-of-the-art and limited uptake in routine clinical practice motivates the need for clinical studies evaluating model performance in practice. Future work will seek to define the clinical acceptability of the segmented organs

XSL•FO
RenderX

at risk produced by our models and estimate the time saving that could be achieved during the radiotherapy planning workflow in a real-world setting.

A number of other study limitations should be addressed in future studies. First, we included only planning CT scans because magnetic resonance imaging and positron emission tomography scans were not routinely performed for all patients in the UCLH data set. Some organ at risk classes, such as optic chiasm, require co-registration with MR images for optimal delineation, and access to additional imaging has been shown to improve the delineation of optic nerves [29]. As a result, certain organ at risk classes were deliberately excluded from this CT-based project and will be addressed in future work that will incorporate magnetic resonance imaging scans. A second limitation is with regard to the classes of organs at risk in this study. Although we presented one of the largest sets of reported organs at risk in the literature [44,97,98], some omissions occurred (eg, oral cavity) owing to an insufficient number of examples in the training data that conformed to a standard international protocol. The number of oncologists used in the creation of our ground truth may not have fully captured the variability in organ at risk segmentation or may have been biased toward a particular interpretation of the Brouwer Atlas used as our segmentation protocol. Even in an organ as simple as the spinal cord that is traditionally reliably outlined by auto-segmentation algorithms, there is ambiguity between the inclusion of, for example, the nerve roots. Such variation may widen the thresholds of acceptable deviation in favor of the model, despite a consistent protocol. Future studies will address these deficits alongside time-consuming lymph node segmentation.

Finally, neither of the test sets used in this study included the patients' protected-characteristic status. This is a significant limitation, as it prevents the study of intersectional fairness.

Conclusions

In conclusion, we demonstrated that deep learning can achieve human expert–level performance in the segmentation of head and neck organs at risk in radiotherapy planning CT scans, using a clinically applicable performance metric designed for this clinical scenario. We provided evidence of the generalizability of this model by testing it on patients from different geographies, demographics, and scanning protocols. This segmentation algorithm was performed with similar accuracy compared with experts and has the potential to improve the speed, efficiency, and consistency of radiotherapy workflows, with an expected positive influence on patient outcomes. Future work will investigate the impact of our segmentation algorithm in clinical practice.

Acknowledgments

The codebase for the deep learning framework makes use of proprietary components, and we are unable to publicly release this code. However, all experiments and implementation details are described in detail in the Methods section to allow independent replication with nonproprietary libraries. The surface DSC performance metric code is available on the internet [99].

The clinical data used for training and validation sets were collected and deidentified at the UCLH NHS Foundation Trust. The data were used for both local and national permissions. They are not publicly available, and restrictions apply to their use. The data, or a subset, may be available from the UCLH NHS Foundation Trust, subject to local and national ethical approvals. The released test or validation set data were collected from two data sets hosted on TCIA. The subset used, along with the ground truth segmentations added, is available on the internet [100].

The authors thank the patients treated at UCLH whose scans were used in this work, A Zisserman, D King, D Barrett, V Cornelius, C Beltran, J Cornebise, R Sharma, J Ashburner, J Good, and N Haji for discussions, M Kosmin for his review of the published literature, J Adler for discussion and review of the manuscript, A Warry, U Johnson, V Rompokos, and the rest of the UCLH Radiotherapy Physics team for work on the data collection, R West for work on the visuals, C Game, D Mitchell, and M Johnson for infrastructure and systems administration, A Paine at Softwire for engineering support at UCLH, A Kitchener and the UCLH Information Governance team for support, J Besley and M Bawn for legal assistance, K Ayoub, K Sullivan, and R Ahmed for initiating and supporting the collaboration, the DeepMind Radiographer Consortium made up of B Garie, Y McQuinlan, K Hampton, S Ireland, K Fuller, H Frank, C Tully, A Jones, and L Turner, and the rest of the DeepMind team for their support, ideas, and encouragement. GR and HM were supported by University College London and the National Institute for Health Research UCLH Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Authors' Contributions

MS, TB, OR, JRL, RM, HM, SAM, DD, CC, and COH initiated the project. SB, RM, DC, CB, and DD, CC, and JRL created the data sets. SB, SN, JDF, AZ, YP, COH, HA, and OR contributed to software engineering. SN, JDF, BRP, and OR designed the model architectures. BG, YMQ, SI, KH and KF manually segmented the images. RM, DC, CB, DD, SAM, HM, GR, COH, AK, and JRL contributed clinical expertise. CM, JRL, TB, SAM, KS, and OR managed the project. COH, CK, ML, JRL, SN, SB, JDF, HM, GR, and OR wrote the paper.

XSL•FO
RenderX

Conflicts of Interest

GR, HM, CK, COH, and DC were paid contractors of DeepMind and Google Health.

Multimedia Appendix 1
Additional Tables S1-S8 and Figures S1-S7 show further visual examples of model outputs, performance metrics and detailed comparisons to previously published works.
[PDF File (Adobe PDF File), 10937 KB - jmir_v23i7e26151_app1.pdf]

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69-90 [FREE Full text] [doi: 10.3322/caac.20107] [Medline: 21296855]
2. Head and neck cancers incidence statistics. Cancer Research UK. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/incidence#heading-Two [accessed 2018-02-08]
3. NCIN data briefing: potentially HPV-related head and neck cancers. National Cancer Intelligence Network. URL: http://www.ncin.org.uk/publications/data_briefings/potentially_hpv_related_head_and_neck_cancers [accessed 2021-05-17]
4. Profile of head and neck cancers in England: incidence, mortality and survival. Oxford Cancer Intelligence Unit. 2010. URL: http://www.ncin.org.uk/view?rid=69 [accessed 2021-05-17]
5. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer 2011 Dec 06;105 Suppl 2:77-81 [FREE Full text] [doi: 10.1038/bjc.2011.489] [Medline: 22158327]
6. Jensen K, Lambertsen K, Grau C. Late swallowing dysfunction and dysphagia after radiotherapy for pharynx cancer: frequency, intensity and correlation with dose and volume parameters. Radiother Oncol 2007 Oct;85(1):74-82. [doi: 10.1016/j.radonc.2007.06.004] [Medline: 17673322]
7. Dirix P, Abbeel S, Vanstraelen B, Hermans R, Nuyts S. Dysphagia after chemoradiotherapy for head-and-neck squamous cell carcinoma: dose-effect relationships for the swallowing structures. Int J Radiat Oncol Biol Phys 2009 Oct 01;75(2):385-392 [FREE Full text] [doi: 10.1016/j.ijrobp.2008.11.041] [Medline: 19553033]
8. Caudell JJ, Schaner PE, Desmond RA, Meredith RF, Spencer SA, Bonner JA. Dosimetric factors associated with long-term dysphagia after definitive radiotherapy for squamous cell carcinoma of the head and neck. Int J Radiat Oncol Biol Phys 2010 Feb 01;76(2):403-409. [doi: 10.1016/j.ijrobp.2009.02.017] [Medline: 19467801]
9. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011 Feb;12(2):127-136. [doi: 10.1016/s1470-2045(10)70290-4]
10. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2012 Jan 01;82(1):368-378. [doi: 10.1016/j.ijrobp.2010.10.019] [Medline: 21123004]
11. Voet PW, Dirkx ML, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJ. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiother Oncol 2011 Mar;98(3):373-377. [doi: 10.1016/j.radonc.2010.11.017] [Medline: 21269714]
12. Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010 Jul 01;77(3):950-958 [FREE Full text] [doi: 10.1016/j.ijrobp.2009.09.062] [Medline: 20378266]
13. Chen Z, King W, Pearcey R, Kerba M, Mackillop WJ. The relationship between waiting time for radiotherapy and clinical outcomes: a systematic review of the literature. Radiother Oncol 2008 Apr;87(1):3-16. [doi: 10.1016/j.radonc.2007.11.016] [Medline: 18160158]
14. Mikeljevic JS, Haward R, Johnston C, Crellin A, Dodwell D, Jones A, et al. Trends in postoperative radiotherapy delay and the effect on survival in breast cancer patients treated with conservation surgery. Br J Cancer 2004 Apr 05;90(7):1343-1348 [FREE Full text] [doi: 10.1038/sj.bjc.6601693] [Medline: 15054452]
15. The NHS Cancer Plan and the new NHS, Chapter 5. National Health Service. 2004. URL: http://www.wales.nhs.uk/technologymls/english/resources/pdf/cancer_nsf.pdf [accessed 2021-05-17]
16. Round C, Williams M, Mee T, Kirkby N, Cooper T, Hoskin P, et al. Radiotherapy demand and activity in England 2006-2020. Clin Oncol (R Coll Radiol) 2013 Sep;25(9):522-530. [doi: 10.1016/j.clon.2013.05.005] [Medline: 23768454]
17. Rosenblatt E, Zubizarreta E. Radiotherapy in cancer care: facing the global challenge. International Atomic Energy Agency. 2017. URL: https://www-pub.iaea.org/MTCD/Publications/PDF/P1638_web.pdf [accessed 2021-05-17]
18. Veiga C, McClelland J, Moinuddin S, Lourenço A, Ricketts K, Annkah J, et al. Toward adaptive radiotherapy for head and neck patients: feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. Med Phys 2014 Mar 19;41(3):031703 [FREE Full text] [doi: 10.1118/1.4864240] [Medline: 24593707]
19. Daisne J, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. Radiat Oncol 2013 Jun 26;8(1):154. [doi: 10.1186/1748-717x-8-154]

XSL•FO
RenderX

20. Fortunati V, Verhaart RF, van der Lijn F, Niessen WJ, Veenland JF, Paulides MM, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. Med Phys 2013 Jul 20;40(7):071905. [doi: 10.1118/1.4810971] [Medline: 23822442]
21. Hoang Duc AK, Eminowicz G, Mendes R, Wong S, McClelland J, Modat M, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. Med Phys 2015 Sep 05;42(9):5027-5034 [FREE Full text] [doi: 10.1118/1.4927567] [Medline: 26328953]
22. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiat Oncol 2014;9(1):173. [doi: 10.1186/1748-717x-9-173]
23. Walker GV, Awan M, Tao R, Koay EJ, Boehling NS, Grant JD, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. Radiother Oncol 2014 Sep;112(3):321-325 [FREE Full text] [doi: 10.1016/j.radonc.2014.08.028] [Medline: 25216572]
24. Gacha SJ, León SA. Segmentation of mandibles in computer tomography volumes of patients with foam cells carcinoma. In: Proceedings of the IX International Seminar of Biomedical Engineering (SIB). 2018 Presented at: IX International Seminar of Biomedical Engineering (SIB); May 16-18, 2018; Bogota, Colombia. [doi: 10.1109/SIB.2018.8467732]
25. Raudaschl PF, Zaffino P, Sharp GC, Spadea MF, Chen A, Dawant BM, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. Med Phys 2017 May 21;44(5):2020-2036 [FREE Full text] [doi: 10.1002/mp.12197] [Medline: 28273355]
26. Wu X, Udupa JK, Tong Y, Odhner D, Pednekar GV, Simone CB, et al. AAR-RT - A system for auto-contouring organs at risk on CT images for radiation therapy planning: Principles, design, and large-scale evaluation on head-and-neck and thoracic cancer cases. Med Image Anal 2019 May;54:45-62 [FREE Full text] [doi: 10.1016/j.media.2019.01.008] [Medline: 30831357]
27. Fritscher K, Raudaschl P, Zaffino P, Spadea MF, Sharp GC, Schubert R. Deep neural networks for fast segmentation of 3D medical images. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Switzerland: Springer; 2016:158-165.
28. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys 2017 Feb 16;44(2):547-557 [FREE Full text] [doi: 10.1002/mp.12045] [Medline: 28205307]
29. Močnik D, Ibragimov B, Xing L, Strojan P, Likar B, Pernuš F, et al. Segmentation of parotid glands from registered CT and MR images. Phys Med 2018 Aug;52:33-41 [FREE Full text] [doi: 10.1016/j.ejmp.2018.06.012] [Medline: 30139607]
30. Ren X, Xiang L, Nie D, Shao Y, Zhang H, Shen D, et al. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. Med Phys 2018 May 23;45(5):2063-2075 [FREE Full text] [doi: 10.1002/mp.12837] [Medline: 29480928]
31. Zhong T, Huang X, Tang F, Liang S, Deng X, Zhang Y. Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma. Med Phys 2019 Sep 16;46(12):5602-5611 [FREE Full text] [doi: 10.1002/mp.13825] [Medline: 31529501]
32. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Switzerland: Springer; 2015:234-241.
33. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018 Sep 13;24(9):1342-1350. [doi: 10.1038/s41591-018-0107-6] [Medline: 30104768]
34. Hänsch A, Schwier M, Gass T, Morgas T, Haas B, Klein J, et al. Comparison of different deep learning approaches for parotid gland segmentation from CT images. Proc. SPIE 10575, Med Imag 2018: Comp-Aid Diag 2018:1057519 [FREE Full text] [doi: 10.1117/12.2292962]
35. Zhu W, Huang Y, Tang H, Qian Z, Du N, Fan W, et al. AnatomyNet: Deep 3D Squeeze-and-excitation U-Nets for fast and fully automated whole-volume anatomical segmentation. BioRxiv 2018:A. [doi: 10.1101/392969]
36. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys 2018 Oct 19;45(10):4558-4567 [FREE Full text] [doi: 10.1002/mp.13147] [Medline: 30136285]
37. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. Eur Radiol 2019 Apr 9;29(4):1961-1967 [FREE Full text] [doi: 10.1007/s00330-018-5748-9] [Medline: 30302589]
38. Willems S, Crijns W, Saint-Esteven AL, Veen JV, Robben D, Depuydt T, et al. Clinical implementation of DeepVoxNet for auto-delineation of organs at risk in head and neck cancer patients in radiotherapy. In: Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Switzerland: Springer; 2018:223-232.
39. Kodym O, Španěl M, Herout A. Segmentation of head and neck organs at risk using CNN with batch dice loss. arXiv.org: Computer Science - Computer Vision and Pattern Recognition. 2018. URL: https://arxiv.org/abs/1812.02427 [accessed 2021-05-17]
40. Wang Y, Zhao L, Wang M, Song Z. Organ at risk segmentation in head and neck CT images using a two-stage segmentation framework based on 3D U-Net. IEEE Access 2019;7:144591-144602 [FREE Full text] [doi: 10.1109/access.2019.2944958]

XSL•FO
RenderX

41. Men K, Geng H, Cheng C, Zhong H, Huang M, Fan Y, et al. Technical Note: More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades. Med Phys 2019 Jan 07;46(1):286-292 [FREE Full text] [doi: 10.1002/mp.13296] [Medline: 30450825]
42. Tappeiner E, Pröll S, Hönig M, Raudaschl PF, Zaffino P, Spadea MF, et al. Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach. Int J Comput Assist Radiol Surg 2019 May 7;14(5):745-754 [FREE Full text] [doi: 10.1007/s11548-019-01922-4] [Medline: 30847761]
43. Rhee DJ, Cardenas CE, Elhalawani H, McCarroll R, Zhang L, Yang J, et al. Automatic detection of contouring errors using convolutional neural networks. Med Phys 2019 Nov 26;46(11):5086-5097 [FREE Full text] [doi: 10.1002/mp.13814] [Medline: 31505046]
44. Tang H, Chen X, Liu Y, Lu Z, You J, Yang M, et al. Clinically applicable deep learning framework for organs at risk delineation in CT images. Nat Mach Intell 2019 Sep 30;1(10):480-491 [FREE Full text] [doi: 10.1038/s42256-019-0099-z]
45. van Rooij W, Dahele M, Brandao HR, Delaney AR, Slotman BJ, Verbakel WF. Deep learning-based delineation of head and neck organs at risk: geometric and dosimetric evaluation. Int J Radiat Oncol Biol Phys 2019 Jul 01;104(3):677-684 [FREE Full text] [doi: 10.1016/j.ijrobp.2019.02.040] [Medline: 30836167]
46. Gou S, Tong N, Qi S, Yang S, Chin R, Sheng K. Self-channel-and-spatial-attention neural network for automated multi-organ segmentation on head and neck CT images. Phys Med Biol 2020 Dec 11;65(24):245034 [FREE Full text] [doi: 10.1088/1361-6560/ab79c3] [Medline: 32097892]
47. Mak RH, Endres MG, Paik JH, Sergeev RA, Aerts H, Williams CL, et al. Use of crowd innovation to develop an artificial intelligence-based solution for radiation therapy targeting. JAMA Oncol 2019 May 01;5(5):654-661 [FREE Full text] [doi: 10.1001/jamaoncol.2019.0159] [Medline: 30998808]
48. Head-neck cetuximab. The Cancer Imaging Archive (TCIA). URL: https://wiki.cancerimagingarchive.net/display/Public/Head-Neck+Cetuximab [accessed 2021-05-17]
49. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013 Dec 25;26(6):1045-1057 [FREE Full text] [doi: 10.1007/s10278-013-9622-7] [Medline: 23884657]
50. Zuley ML, Jarosz R, Kirk S, Colen R, Garcia K, Aredes ND. Radiology data from the cancer genome atlas head-neck squamous cell carcinoma [TCGA-HNSC] collection. The Cancer Imaging Archive (TCIA) 2020:A [FREE Full text] [doi: 10.7937/K9/TCIA.2016.LXKQ47MS]
51. Brouwer CL, Steenbakkers RJ, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol 2015 Oct;117(1):83-90. [doi: 10.1016/j.radonc.2015.07.041]
52. Felzenszwalb PF, Huttenlocher DP. Distance transforms of sampled functions. Theory Comput 2012;8(1):415-428. [doi: 10.4086/toc.2012.v008a019]
53. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv.org : Computer Science - Machine Learning. 2014. URL: http://arxiv.org/abs/1412.6980 [accessed 2021-05-17]
54. Wu Z, Shen C, van den Hengel A. Bridging category-level and instance-level semantic image segmentation. arXiv.org : Computer Science - Computer Vision and Pattern Recognition. 2016. URL: http://arxiv.org/abs/1605.06885v1 [accessed 2021-05-17]
55. Lorensen WE, Cline HE. Marching cubes: a high resolution 3D surface construction algorithm. SIGGRAPH Comput Graph 1987 Aug;21(4):163-169. [doi: 10.1145/37402.37422]
56. Wang Z, Wei L, Wang L, Gao Y, Chen W, Shen D. Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning. IEEE Trans Image Process 2018 Feb;27(2):923-937. [doi: 10.1109/tip.2017.2768621]
57. Torosdagli N, Liberton DK, Verma P, Sincan M, Lee J, Pattanaik S, et al. Robust and fully automated segmentation of mandible from CT scans. In: Proceedings of the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). 2017 Presented at: IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); April 18-21, 2017; Melbourne, VIC, Australia. [doi: 10.1109/isbi.2017.7950734]
58. Liang S, Thung K, Nie D, Zhang Y, Shen D. Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images. IEEE Trans Med Imaging 2020 Sep;39(9):2794-2805 [FREE Full text] [doi: 10.1109/tmi.2020.2975853]
59. Qiu B, Guo J, Kraeima J, Glas HH, Borra RJ, Witjes MJ, et al. Recurrent convolutional neural networks for mandible segmentation from computed tomography. arXiv.org: Electrical Engineering and Systems Science - Image and Video Processing. 2020. URL: https://arxiv.org/abs/2003.06486 [accessed 2021-05-27]
60. Sun S, Liu Y, Bai N, Tang H, Chen X, Huang Q, et al. Attentionanatomy: a unified framework for whole-body organs at risk segmentation using multiple partially annotated datasets. In: Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020 Presented at: IEEE 17th International Symposium on Biomedical Imaging (ISBI); April 3-7, 2020; Iowa City, IA, USA p. A. [doi: 10.1109/isbi45749.2020.9098588]
61. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers RJ, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. Radiother Oncol 2020 Jan;142:115-123 [FREE Full text] [doi: 10.1016/j.radonc.2019.09.022] [Medline: 31653573]

XSL•FO
RenderX

62. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiother Oncol 2020 Mar;144:152-158 [FREE Full text] [doi: 10.1016/j.radonc.2019.10.019] [Medline: 31812930]
63. Chan JW, Kearney V, Haaf S, Wu S, Bogdanov M, Reddick M, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. Med Phys 2019 May 04;46(5):2204-2213 [FREE Full text] [doi: 10.1002/mp.13495] [Medline: 30887523]
64. Gao Y, Huang R, Chen M, Wang Z, Deng J, Chen Y, et al. FocusNet: imbalanced large and small organ segmentation with an end-to-end deep neural network for head and neck CT images. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Switzerland: Springer; 2019:829-838.
65. Jiang J, Sharif E, Um H, Berry S, Veeraraghavan H. Local block-wise self attention for normal organ segmentation. arXiv.org: Computer Science - Computer Vision and Pattern Recognition. 2019. URL: https://arxiv.org/abs/1909.05054 [accessed 2021-05-16]
66. Lei W, Wang H, Gu R, Zhang S, Wang G. DeepIGeoS-V2: Deep interactive segmentation of multiple organs from head and neck images with lightweight CNNs. In: Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. Switzerland: Springer; 2019:61-69.
67. Sun Y, Shi H, Zhang S, Wang P, Zhao W, Zhou X, et al. Accurate and rapid CT image segmentation of the eyes and surrounding organs for precise radiotherapy. Med Phys 2019 May 22;46(5):2214-2222 [FREE Full text] [doi: 10.1002/mp.13463] [Medline: 30815885]
68. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. Med Phys 2019 Jun 06;46(6):2669-2682 [FREE Full text] [doi: 10.1002/mp.13553] [Medline: 31002188]
69. Xue Y, Tang H, Qiao Z, Gong G, Yin Y, Qian Z, et al. Shape-aware organ segmentation by predicting signed distance maps. arXiv.org: Computer Science - Computer Vision and Pattern Recognition 2020 Apr 03;34(07):12565-12572 [FREE Full text] [doi: 10.1609/aaai.v34i07.6946]
70. Nikolov S, Blackwell S, Mendes R, Fauw JD, Meyer C, Hughes C, DeepMind Radiographer Consortium, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. arXiv.org: Computer Science - Computer Vision and Pattern Recognition. 2018. URL: https://arxiv.org/abs/1809.04430v1 [accessed 2021-05-16]
71. Fritscher KD, Peroni M, Zaffino P, Spadea MF, Schubert R, Sharp G. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. Med Phys 2014 May 24;41(5):051910 [FREE Full text] [doi: 10.1118/1.4871623] [Medline: 24784389]
72. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach. Med Phys 2011 Nov 26;38(11):6160-6170 [FREE Full text] [doi: 10.1118/1.3654160] [Medline: 22047381]
73. Tam CM, Yang X, Tian S, Jiang X, Beitler JJ, Li S. Automated delineation of organs-at-risk in head and neck CT images using multi-output support vector regression. In: Proceedings of the SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. 2018 Presented at: SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging; March 12, 2018; Houston, Texas, United States p. -. [doi: 10.1117/12.2292556]
74. Wang Z, Liu X, Chen W. Head and neck CT atlases alignment based on anatomical priors constraint. J Med Imaging Health Infor 2019 Dec 01;9(9):2004-2011 [FREE Full text] [doi: 10.1166/jmihi.2019.2844]
75. Ayyalusamy A, Vellaiyan S, Subramanian S, Ilamurugu A, Satpathy S, Nauman M, et al. Auto-segmentation of head and neck organs at risk in radiotherapy and its dependence on anatomic similarity. Radiat Oncol J 2019 Jun;37(2):134-142 [FREE Full text] [doi: 10.3857/roj.2019.00038] [Medline: 31266293]
76. Haq R, Berry SL, Deasy JO, Hunt M, Veeraraghavan H. Dynamic multiatlas selection-based consensus segmentation of head and neck structures from CT images. Med Phys 2019 Dec 31;46(12):5612-5622 [FREE Full text] [doi: 10.1002/mp.13854] [Medline: 31587300]
77. McCarroll RE, Beadle BM, Balter PA, Burger H, Cardenas CE, Dalvie S, et al. Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: a step toward automated radiation treatment planning for low- and middle-income countries. J Glob Oncol 2018 Dec(4):1-11 [FREE Full text] [doi: 10.1200/jgo.18.00055]
78. Liu Q, Qin A, Liang J, Yan D. Evaluation of atlas-based auto-segmentation and deformable propagation of organs-at-risk for head-and-neck adaptive radiotherapy. Recent Pat Top Imaging 2016 May 24;5(2):79-87 [FREE Full text] [doi: 10.2174/2451827105999160415123925]
79. Tao C, Yi J, Chen N, Ren W, Cheng J, Tung S, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. Radiother Oncol 2015 Jun;115(3):407-411 [FREE Full text] [doi: 10.1016/j.radonc.2015.05.012] [Medline: 26025546]
80. Wachinger C, Fritscher K, Sharp G, Golland P. Contour-driven atlas-based segmentation. IEEE Trans Med Imaging 2015 Dec;34(12):2492-2505 [FREE Full text] [doi: 10.1109/tmi.2015.2442753]

XSL•FO
RenderX

81. Zhu M, Bzdusek K, Brink C, Eriksen JG, Hansen O, Jensen HA, et al. Multi-institutional quantitative evaluation and clinical validation of Smart Probabilistic Image Contouring Engine (SPICE) autosegmentation of target structures and normal tissues on computer tomography images in the head and neck, thorax, liver, and male pelvis areas. Int J Radiat Oncol Biol Phys 2013 Nov 15;87(4):809-816 [FREE Full text] [doi: 10.1016/j.ijrobp.2013.08.007] [Medline: 24138920]
82. Teguh DN, Levendag PC, Voet PW, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys 2011 Nov 15;81(4):950-957. [doi: 10.1016/j.ijrobp.2010.07.009] [Medline: 20932664]
83. Han X, Hibbard LS, O'Connell NP, Willcut V. Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. ResearchGate. 2011. URL: https://www.researchgate.net/profile/Lyndon-Hibbard/publication/228519091_Automatic_Segmentation_of_Parotids_in_Head_and_Neck_CT_Images_using_Multi-atlas_Fusion/links/0deec516d54dfccb97000000/Automatic-Segmentation-of-Parotids-in-Head-and-Neck-CT-Images-using-Multi-atlas-Fusion.pdf [accessed 2021-05-27]
84. Sims R, Isambert A, Grégoire V, Bidault F, Fresco L, Sage J, et al. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. Radiother Oncol 2009 Dec;93(3):474-478. [doi: 10.1016/j.radonc.2009.08.013] [Medline: 19758720]
85. Han X, Hoogeman MS, Levendag PC, Hibbard LS, Teguh DN, Voet P, et al. Atlas-based auto-segmentation of head and neck CT images. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. Berlin, Heidelberg: Springer; 2008:434-441.
86. Hoogeman M, Han X, Teguh D, Voet P, Nowak P, Wolf T, et al. Atlas-based auto-segmentation of CT images in head and neck cancer: what is the best approach? Int J Radiat Oncol Biol Phys 2008 Sep;72(1):591 [FREE Full text] [doi: 10.1016/j.ijrobp.2008.06.196]
87. Huang C, Badiei M, Seo H, Ma M, Liang X, Capaldi D, et al. Atlas based segmentations via semi-supervised diffeomorphic registrations. arXiv.org: Computer Science - Computer Vision and Pattern Recognition. 2019. URL: https://arxiv.org/abs/1911.10417 [accessed 2021-05-16]
88. Hardcastle N, Tomé WA, Cannon DM, Brouwer CL, Wittendorp PW, Dogan N, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. Radiat Oncol 2012 Jun 15;7(1):90 [FREE Full text] [doi: 10.1186/1748-717x-7-90]
89. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiat Oncol 2012;7(1):160 [FREE Full text] [doi: 10.1186/1748-717x-7-160]
90. Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. Int J Radiat Oncol Biol Phys 2007 Jun 01;68(2):522-530 [FREE Full text] [doi: 10.1016/j.ijrobp.2007.01.038] [Medline: 17418960]
91. Dice LR. Measures of the amount of ecologic association between species. Ecol Soc Am 1945;26(3):297-302. [doi: 10.2307/1932409]
92. Hong T, Tome W, Chappel R, Harari P. Variations in target delineation for head and neck IMRT: an international multi-institutional study. Int J Radiat Oncol Biol Phys 2004 Sep;60:157-158 [FREE Full text] [doi: 10.1016/s0360-3016(04)01130-7]
93. Wuthrick EJ, Zhang Q, Machtay M, Rosenthal DI, Nguyen-Tan PF, Fortin A, et al. Institutional clinical trial accrual volume and survival of patients with head and neck cancer. J Clin Oncol 2015 Jan 10;33(2):156-164 [FREE Full text] [doi: 10.1200/jco.2014.56.5218]
94. Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. Phys Imaging Radiat Oncol 2020 Jan;13:1-6 [FREE Full text] [doi: 10.1016/j.phro.2019.12.001] [Medline: 33458300]
95. Kiser KJ, Barman A, Stieb S, Fuller CD, Giancardo L. Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. medRxiv 2020. [doi: 10.1101/2020.05.14.20102103]
96. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys 2014 May 24;41(5):050902 [FREE Full text] [doi: 10.1118/1.4871620] [Medline: 24784366]
97. Kosmin M, Ledsam J, Romera-Paredes B, Mendes R, Moinuddin S, de Souza D, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. Radiother Oncol 2019 Jun;135:130-140 [FREE Full text] [doi: 10.1016/j.radonc.2019.03.004] [Medline: 31015159]
98. Guo D, Jin D, Zhu Z, Ho TY, Harrison HP, Chao CH, et al. Organ at risk segmentation for head and neck cancer using stratified learning and neural architecture search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 13-19, 2020; Seattle, WA, USA. [doi: 10.1109/cvpr42600.2020.00428]
99. Surface distance. DeepMind. URL: https://github.com/deepmind/surface-distance [accessed 2021-05-27]
100. TCIA CT scan dataset. DeepMind. URL: https://github.com/deepmind/tcia-ct-scan-dataset [accessed 2021-05-27]

XSL•FO
RenderX

Abbreviations

CT: computed tomography
DSC: Dice similarity coefficient
NHS: National Health Service
PDDCA: public domain database for computational anatomy
TCGA-HNSC: The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma
TCIA: The Cancer Imaging Archive
UCLH: University College London Hospitals

Edited by R Kukafka; submitted 30.11.20; peer-reviewed by JA Benítez-Andrades, R Vilela; comments to author 11.01.21; revised version received 10.02.21; accepted 30.04.21; published 12.07.21.

Please cite as:
Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, Patel Y, Meyer C, Askham H, Romera-Paredes B, Kelly C, Karthikesalingam A, Chu C, Carnell D, Boon C, D'Souza D, Moinuddin SA, Garie B, McQuinlan Y, Ireland S, Hampton K, Fuller K, Montgomery H, Rees G, Suleyman M, Back T, Hughes CO, Ledsam JR, Ronneberger O
Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study
J Med Internet Res 2021;23(7):e26151
URL: https://www.jmir.org/2021/7/e26151
doi:10.2196/26151
PMID:34255661

©Stanislav Nikolov, Sam Blackwell, Alexei Zverovitch, Ruheena Mendes, Michelle Livne, Jeffrey De Fauw, Yojan Patel, Clemens Meyer, Harry Askham, Bernadino Romera-Paredes, Christopher Kelly, Alan Karthikesalingam, Carlton Chu, Dawn Carnell, Cheng Boon, Derek D'Souza, Syed Ali Moinuddin, Bethany Garie, Yasmin McQuinlan, Sarah Ireland, Kiarna Hampton, Krystle Fuller, Hugh Montgomery, Geraint Rees, Mustafa Suleyman, Trevor Back, Cían Owen Hughes, Joseph R Ledsam, Olaf Ronneberger. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of an Anomaly Detection Model to Screen for Ocular Diseases Using Color Retinal Fundus Images: Design and Evaluation Study

Yong Han[1,2], PhD; Weiming Li[1,2], PhD; Mengmeng Liu[1,2], MS; Zhiyuan Wu[1,2], PhD; Feng Zhang[1,2], PhD; Xiangtong Liu[1,2], PhD; Lixin Tao[1,2], PhD; Xia Li[3], PhD; Xiuhua Guo[1,2], PhD

[1]Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, China

[2]Beijing Municipal Key Laboratory of Clinical Epidemiology, Capital Medical University, Beijing, China

[3]Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia

Corresponding Author:
Xiuhua Guo, PhD
Department of Epidemiology and Health Statistics, School of Public Health
Capital Medical University
No 10 Xitoutiao, You'anmen Wai, Fengtai District
Beijing, 1000069
China
Phone: 86 13661283546
Email: statguo@ccmu.edu.cn

Abstract

Background: The supervised deep learning approach provides state-of-the-art performance in a variety of fundus image classification tasks, but it is not applicable for screening tasks with numerous or unknown disease types. The unsupervised anomaly detection (AD) approach, which needs only normal samples to develop a model, may be a workable and cost-saving method of screening for ocular diseases.

Objective: This study aimed to develop and evaluate an AD model for detecting ocular diseases on the basis of color fundus images.

Methods: A generative adversarial network–based AD method for detecting possible ocular diseases was developed and evaluated using 90,499 retinal fundus images derived from 4 large-scale real-world data sets. Four other independent external test sets were used for external testing and further analysis of the model's performance in detecting 6 common ocular diseases (diabetic retinopathy [DR], glaucoma, cataract, age-related macular degeneration, hypertensive retinopathy [HR], and myopia), DR of different severity levels, and 36 categories of abnormal fundus images. The area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity of the model's performance were calculated and presented.

Results: Our model achieved an AUC of 0.896 with 82.69% sensitivity and 82.63% specificity in detecting abnormal fundus images in the internal test set, and it achieved an AUC of 0.900 with 83.25% sensitivity and 85.19% specificity in 1 external proprietary data set. In the detection of 6 common ocular diseases, the AUCs for DR, glaucoma, cataract, AMD, HR, and myopia were 0.891, 0.916, 0.912, 0.867, 0.895, and 0.961, respectively. Moreover, the AD model had an AUC of 0.868 for detecting any DR, 0.908 for detecting referable DR, and 0.926 for detecting vision-threatening DR.

Conclusions: The AD approach achieved high sensitivity and specificity in detecting ocular diseases on the basis of fundus images, which implies that this model might be an efficient and economical tool for optimizing current clinical pathways for ophthalmologists. Future studies are required to evaluate the practical applicability of the AD approach in ocular disease screening.

(J Med Internet Res 2021;23(7):e27822) doi:10.2196/27822

KEYWORDS

anomaly detection; artificial intelligence; cataract; diabetic retinopathy; disease screening; eye; fundus image; glaucoma; macular degeneration; ocular disease; ophthalmology

XSL•FO
RenderX

Introduction

Globally, approximately 2.2 billion people have vision impairment or blindness, according to the first World Report on Vision issued by the World Health Organization in 2019 [1]. In addition, the growth of the global population and the changes in its age structure are leading to accelerated growth of this number [2]. There is adequate evidence that retinal screening and referral for treatment can prevent avoidable blindness [3]. Advances in fundus imaging are expected to decrease preventable visual morbidity by enabling convenient and timely eye disease screening [4,5]. Color fundus camera imaging is an essential and easy-to-master technique for detecting a variety of eye diseases, such as diabetic retinopathy (DR) [6], age-related macular degeneration (AMD) [7], glaucoma [8], cataracts [9], and myopia [10,11]. However, in most countries, especially in low-income countries or regions with insufficient medical resources, there are not enough highly skilled ophthalmologists engaged in eye screening. Therefore, there is an urgent need to develop a convenient and low-cost auxiliary diagnostic system to screen for ocular diseases.

With the promotion of artificial intelligence (AI) in medicine over the past decade [12,13], AI has proven to be effective and feasible for automatic eye disease screening or diagnosis based on color retinal fundus images. Over the past few years, supervised deep learning approaches have provided state-of-the-art performance in eye disease classification tasks and have achieved excellent sensitivity and specificity in the detection of DR [14-16], AMD [17,18], cataracts [11], glaucoma [15,19], and pathological myopia [20]. However, these models may fail in real-life settings because they are trained on the basis of only one type of fundus disease, when a given fundus image used for inference is neither normal nor the specific ocular disease used in training; in such situations, the model is bound to make an incorrect prediction. A multiclass model that includes all types of ocular diseases may be an option to consider. However, such a model would require large data sets annotated by ophthalmologists, which is laborious and expensive. Furthermore, the gamut of all possible anomalies is not available in most cases owing to the rarity of certain diseases and disorders.

In disease screening, the proportion of normal samples is generally much higher than that of anomalies, which implies that the task is akin to anomaly detection (AD). In AD, a model is developed on the basis of only normal samples to capture the distribution of normality and is then evaluated on both unseen normal and abnormal samples to test their deviation from the distribution [21]. Numerous previous studies have focused on AD in understanding visual scenes [22-25], with a wide range of application domains [26-29]. However, AD has rarely been applied to medical images, where the distinctions between normality and anomaly may be subtler and more variable than those in natural images.

In the field of ophthalmology, Seeböck et al [30] developed the first AD system by training a 1-class support vector machine model unsupervised to identify anomalous regions in optical coherence tomography (OCT) images. Furthermore, a series of generative adversarial network (GAN)–based AD methods were proposed for OCT AD, which demonstrated excellent performance [31-33]. To date, only 1 study has adopted the isolation forest AD algorithm to detect ocular diseases on the basis of small-scale data sets of color retinal fundus images [34]. The area under the receiver operating characteristic curve (AUC) of the model for detecting premature retinopathy and DR were 0.770 and 0.745, respectively, which do yet meet clinical requirements. Hence, we decided to test the prospects of a deep learning–based AD model developed on a large-scale data set of color fundus images.

The primary aim of our study was to develop an AD model based on normal retinal fundus images for the detection of ocular diseases. Our secondary aim was to evaluate the performance of this approach in detecting 6 common ocular diseases, DR of different severity grades, and a variety of fundus abnormalities.

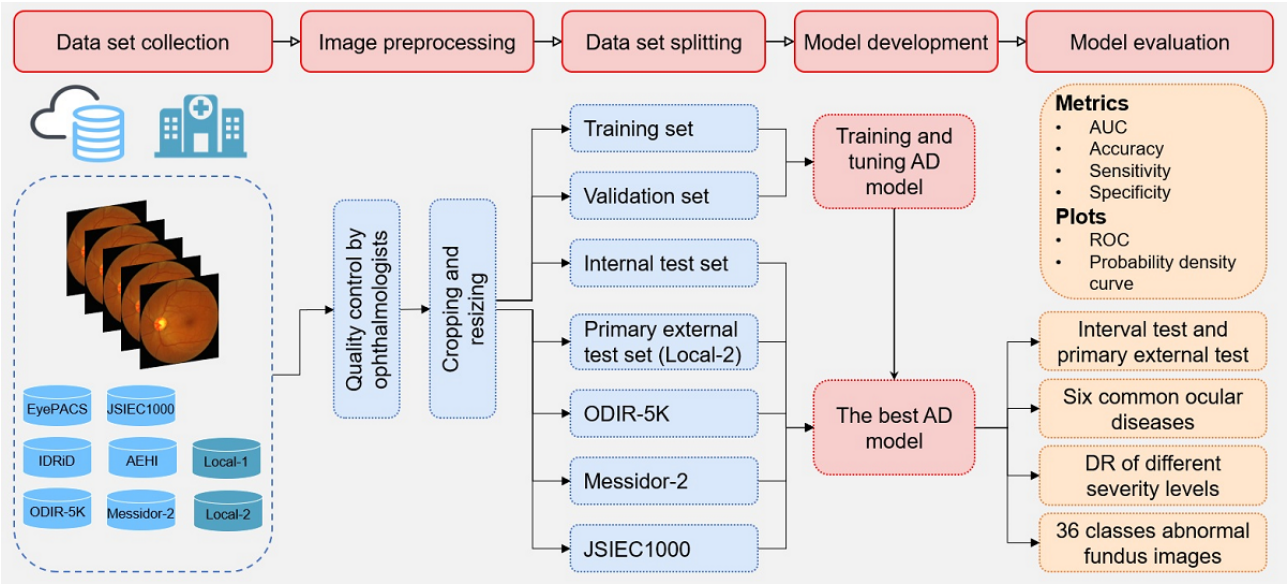
Methods

Methods Overview

Figure 1 represents the workflow for the establishment and verification of the proposed AD method in this study. The first step was to compile fundus images from various sources into a large-scale data set, which contains public data sets available on websites and data sets of our own. The next step was to preprocess the fundus images, including manual removal of fundus images that are not qualified for diagnosis because of substandard quality, as well as cropping of images from different sources to a uniform size. In the third step, the AD model was developed on the training and validation sets. Specifically, the training set was applied to optimize the learnable parameters of the model, whereas the validation set was used to determine the best configuration of the hyperparameters (such as the learning rate, batch size, and momentum) through the random search strategy. The optimal hyperparameters refer to the combination of hyperparameters that can make the AD model achieve the highest AUC on the validation set. In the last step, the final model was first evaluated on internal and primary external test sets, and then we evaluated the model's ability to detect 6 common ocular diseases, different severity levels of DR, and 36 types of fundus abnormal findings or diseases.

XSL•FO
RenderX

Figure 1. Workflow of this study. AD: anomaly detection, AUC: area under the receiver operating characteristic curve, DR: diabetic retinopathy, ODIR: Ocular Disease Intelligent Recognition, and ROC: receiver operating characteristic curve.



Training, Validation, and Test Sets

Eight color fundus image data sets from various sources were collected for model development and evaluation in our study, which are described in Table 1. These fundus images were collected from numerous clinical or health care institutions with diverse models of color fundus cameras in 4 countries (United States, India, China, and France). All fundus images were centered near the macula, while the pupil dilation and field of vision were inconsistent.

We compiled 4 data sets (3 publicly available and 1 proprietary), which comprise 64,351 normal fundus images and 26,148 fundus images with lesions, as training, validation, and internal test sets. Since the model only needs normal fundus images for training, we randomly sampled 60% of normal fundus images as the training set, 20% of normal and 50% of abnormal fundus images as the validation set for hyperparameter tuning of the model during the training phase, and the remaining 20% of normal and 50% of abnormal fundus images as the test set for internal testing of the model's performance.

The model was externally tested using 4 additional fundus image data sets, of which 1 is proprietary and the other 3 are publicly available (Table 1). Ocular Disease Intelligent Recognition (ODIR-5K) contains images of 6 common ocular diseases, namely DR (1334 diagnosable images), glaucoma (456 images), cataract (439 images), AMD (322 images), HR (251 images), and myopia (377 images). The Messidor-2 data set is a collection of DR examinations [35,36], which contains only abnormal images of DR, and each image is rated by a medical expert as 1 of 5 severity levels in accordance with the International Clinical Diabetic Retinopathy Disease Severity Scale [37]. The JSIEC1000 data set contains 36 categories of abnormal retinal fundus images, including some rare diseases such as retinitis pigmentosa, congenital disc abnormality, fundus neoplasm, and Vogt-Koyanagi-Harada disease.

We used these 4 data sets to further evaluate different aspects of model performance. The primary analysis was performed to evaluate the performance of the AD algorithm in detecting all abnormal fundus images in the Local-2 data set. Next, the following subsidiary analyses were performed: (1) the detection of 6 common ocular diseases was analyzed in the ODIR-5K data set, (2) the detection of different severities of DR was evaluated in the Messidor-2 data set, and (3) the detection of 36 abnormal findings or diseases was assessed on the basis of the JSIEC1000 data set.

XSL•FO
RenderX

Table 1. Summary of all data sets used to develop and evaluate the anomaly detection model.

Source data sets	Abbreviation	Race or ethnicity	Cohort	Camera models	Annotators	Number of diagnosable images	
						Normal	Abnormal
Training, validation and internal test sets							
EyePACS program in California (United States)	EyePACS	Various ethnicities	Clinic-based	A variety of cameras	A panel of medical specialists	47,306	23,301
Health Examination Center of Beijing Xiaotangshan Hospital (China)	Local-1	Chinese	Population-based	Topcon TRC-NW100	1 ophthalmologist	14,933	987
Aravind Eye Hospital (India)	AEHI	Indian	Clinic-based	A variety of cameras	1 clinician	1764	1692
Eye Clinic of Sushrusha Hospital (India)	IDRiD	Indian	Clinic-based	Unclear	Medical experts	348	168
External test sets							
Beijing Physical Examination Center (China)	Local-2	Chinese	Population-based	Topcon TRC-NW400	1 ophthalmologist	17,027	1089
More than 400 clinical hospitals in China (China)	ODIR-5K	Chinese	Clinic-based	A variety of cameras	Trained human readers	3492	3179
Ophthalmology department of Brest University Hospital (France)	Messidor-2	French	Clinic-based	Topcon TRC NW6	1 medical expert	1017	731
Joint Shantou International Eye Centre (China)	JSIEC1000	Chinese	Clinic-based	A variety of cameras	Medical experts	54	946

Image Preprocessing

Our data sets were obtained from a wide range of real-world sources, and their characteristics reflect their origins; some images may be out of focus, of inaccurate exposure, or contain artefacts and noise that are not relevant to the diagnosis. For unsupervised AD methods, it is necessary to eliminate these low-quality fundus images to ensure that they are not recognized as abnormal images [38]. Two trained junior ophthalmologists (with 2~3 years of experience) were asked to identify and discard low-quality fundus images that were insufficient to make a reliable diagnosis independently. In case of discordance between these 2 screeners, arbitration was performed by a senior ophthalmologist (with 12 years of experience) to generate a final judgment. The number of diagnosable images in each data set is listed in Table 1.

Moreover, owing to the diverse fundus camera models and settings, the regions captured by the fundus photographs are also markedly heterogeneous, which may cause the model to learn features that are extraneous to disease diagnosis. Hence, we cropped all fundus images to save the same area and standardized the image sizes to a width and height of 800 pixels and 660 pixels, respectively. Multimedia Appendix 1 provides details regarding the pipeline for fundus image preprocessing. We determined the optimal input image resolution for loading into the model through a pilot study, which investigated image resolutions ranging from 32 × 32 pixels to 640 × 640 pixels. The results showed that an image resolution of 256 × 256 pixels achieved the maximum AUC (Multimedia Appendix 2). Hence, we reshaped the fundus images to 256 × 256 pixels when loading them into the AD model for training or inference.

AD Algorithm

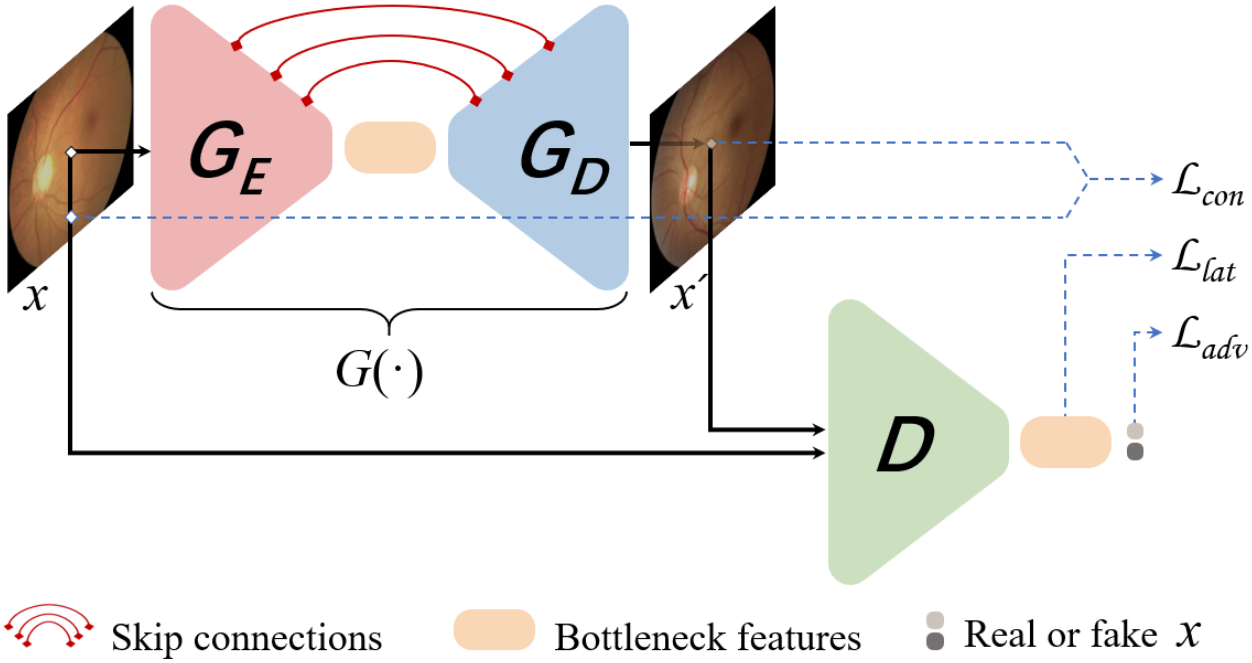
In this study, we adopted an AD method, known as Skip-GANomaly, proposed by Akçay et al [21]. This approach learns representations within both image and latent vector space jointly and achieves state-of-the-art performance both statistically and computationally.

As shown in Figure 2, this algorithm is a GAN network that includes a generator (G) and a discriminator (D) network. The generative network is similar to an automatic encoder, which comprises an encoder (G_E) and a decoder (G_D) network and skip connections between them. In the training phase, the model is trained only on normal samples, and the training objective is to capture the distribution of normal images within not only image space but also latent vector space. To achieve this goal, 3 loss equations are established: adversarial loss (L_{adv}), contextual loss (L_{con}), and latent loss (L_{lat}). The total training objective becomes a weighted sum of the aforementioned losses. In the inference phase, an anomaly score composed of weighted

XSL•FO
RenderX

contextual loss (L_{con}) and latent loss (L_{lat}) was used to detect the anomalies. The training objective would yield minimum anomaly scores for training samples (normal samples) but higher scores for abnormal images. Hence, a higher anomaly score for a given sample x indicates that x is likely abnormal with respect to the distribution of normal samples learned by the AD model from the training set during training [21]. The optimal threshold for the anomaly scores was determined using the Youden index [39]; that is, the critical threshold value that achieved the maximum Youden index was referred to as the optimal threshold.

Figure 2. Overview of the training procedure of the Skip-GANomaly model.



We trained the AD model in an unsupervised manner on the training set (38,611 normal samples) and tuned the hyperparameters of the model on the validation set (containing 12,870 normal and 13,074 abnormal samples). The configuration of the optimal hyperparameters, the optimization objective of the model in the training phase, and the calculation method of anomaly score in the inference phase are detailed in Multimedia Appendix 3.

Statistical Analysis

The metrics AUC, accuracy, sensitivity, and specificity were used to evaluate the performance of the AD algorithm. For the calculation, abnormal fundus images were regarded as positive samples, and normal fundus images were regarded as negative. The binomial exact 95% CI was calculated for the AUC, and the Wilson score was applied to calculate the Wilson 95% CIs for accuracy, sensitivity, and specificity. All statistical analyses were conducted using R (version 3.6.0, The R Foundation).

Results

In total, the AD algorithm for detecting abnormal fundus images was developed using a training set of 38,611 normal fundus images and a validation set comprising 12,870 normal fundus images and 13,074 anomalies. With the optimal hyperparameters derived from model tuning, a final complete model was trained on all 51,481 images, including all normal fundus images in the training set and the validation set. Then, we evaluated the model's performance in detecting abnormal fundus images in the internal and external test sets.

First, we investigated the model's performance on the internal test set (containing 12,870 normal fundus images and 13,074 anomalies) and the external test set Local-2 (containing 17,027 normal fundus images and 1089 anomalies). The algorithm achieved an AUC of 0.896 on the internal test set (Table 2 and Figure 3). The maximum Youden index was 0.6536 (sensitivity=82.69% and specificity=82.63%), which corresponded to the optimal threshold of 2.874×10^{-3}. The model's performance on external test set Local-2 was equal to its performance on the internal test set (Table 2).

XSL•FO
RenderX

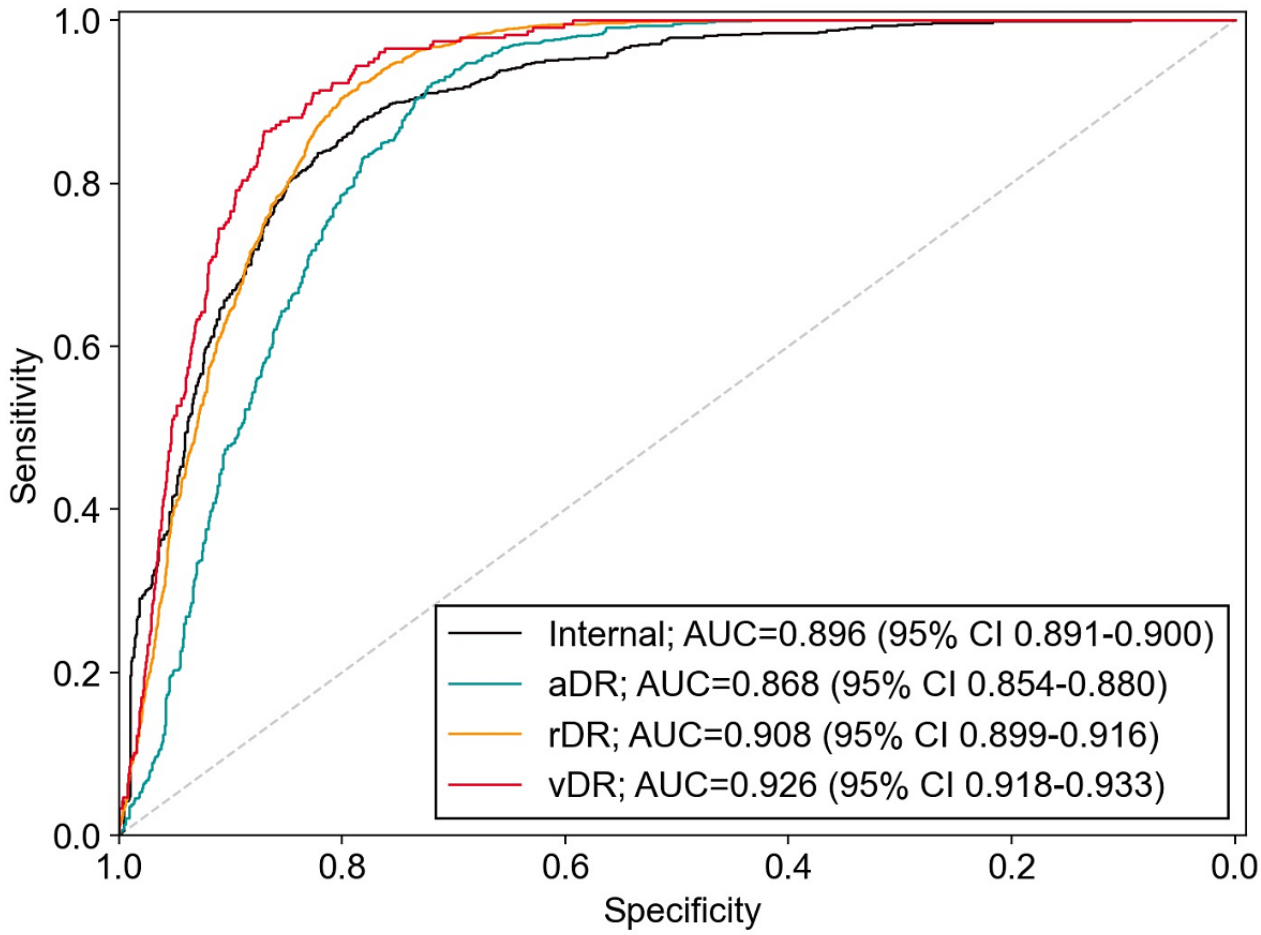
Table 2. The performance of the anomaly detection model in detecting abnormal fundus images in the internal and external test sets.

Test data sets	Area under the receiver operating characteristic curve (95% CI)[a]	Proportion (%) (95% CI)[b]		
		Accuracy	Sensitivity	Specificity
Internal test set	0.896 (0.891- 0.900)	82.67 (82.10-83.22)	82.69 (81.94-83.42)	82.63 (81.74-83.49)
Local-2	0.900 (0.893-0.906)	83.35 (82.63-84.04)	83.25 (82.52-83.97)	85.19 (81.90-88.07)

[a]The binomial exact 95% CI was calculated for each are under the receiver operating characteristic curve.

[b]The Wilson score was applied to calculate the Wilson 95% CI for accuracy, sensitivity, and specificity.

Figure 3. ROC and AUC of the anomaly detection model for detecting abnormalities in the internal test set, as well as detecting aDR, rDR, and vDR in the Messidor-2 data set. “internal” refers to internal test set. aDR diabetic retinopathy of any severity, AUC: area under the receiver operating characteristic curve, rDR: referable diabetic retinopathy, ROC: receiver operating characteristic, vDR vision-threatening diabetic retinopathy.



Next, to explore the model’s capacity to detect different ocular diseases, we evaluated the model’s effectiveness in detecting 6 common types of ocular diseases by using the ODIR-5K data set (Table 3 and Figure 4). The top 3 AUCs were 0.961, 0.916, and 0.912, which corresponded to myopia, glaucoma, and cataract, respectively. AMD had the lowest AUC of 0.867 among all 6 categories. The model had greater than 80% accuracy, sensitivity, and specificity for the detection of all evaluated ocular diseases, except for AMD. Furthermore, we calculated the anomaly score of each fundus image and plotted the distribution of the scores by using probability density curves (Figure 5).

XSL•FO
RenderX

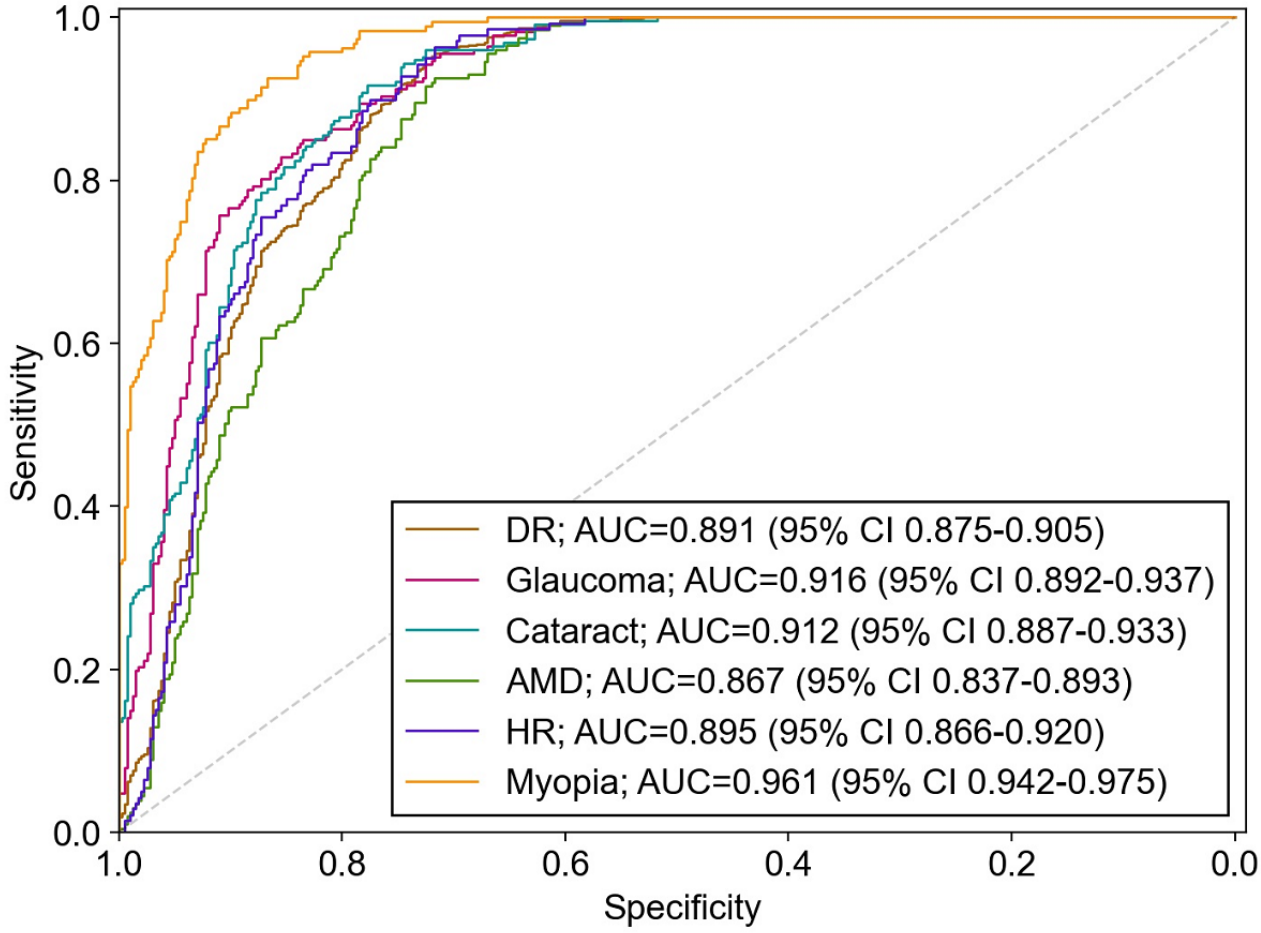
Table 3. The Ocular Disease Intelligent Recognition-5K test set: area under the receiver operating characteristic curve, accuracy, sensitivity, and specificity of the anomaly detection model in detecting 6 ocular diseases.

Ocular diseases	Area under the receiver operating characteristic curve (95% CI)[a]	Proportion (%) (95% CI)[b]		
		Accuracy	Sensitivity	Specificity
All anomalies	0.896 (0.885-0.906)	80.72 (79.39-81.99)	80.69 (79.27-82.03)	81.00 (76.87-84.54)
Diabetic retinopathy	0.891 (0.875-0.905)	80.39 (78.42-82.23)	80.36 (78.09-82.45)	80.50 (76.33-84.08)
Glaucoma	0.916 (0.892-0.937)	83.89 (80.81-86.56)	83.70 (78.34-87.94)	84.00 (80.09-87.27)
Cataract	0.912 (0.887-0.933)	83.44 (80.33-86.14)	83.33 (77.95-87.61)	83.50 (79.55-86.82)
Age-related macular degeneration	0.867 (0.837-0.893)	78.37 (74.90-81.48)	77.61 (71.36-82.83)	78.75 (74.48-82.48)
Hypertensive retinopathy	0.895 (0.866-0.920)	82.37 (78.93-85.36)	81.29 (74.00-86.90)	82.75 (78.74-86.14)
Myopia	0.961 (0.942-0.975)	88.78 (85.97-91.08)	88.83 (83.53-92.58)	88.75 (85.28-91.49)

[a]The binomial exact 95% CI was calculated for each are under the receiver operating characteristic curve.

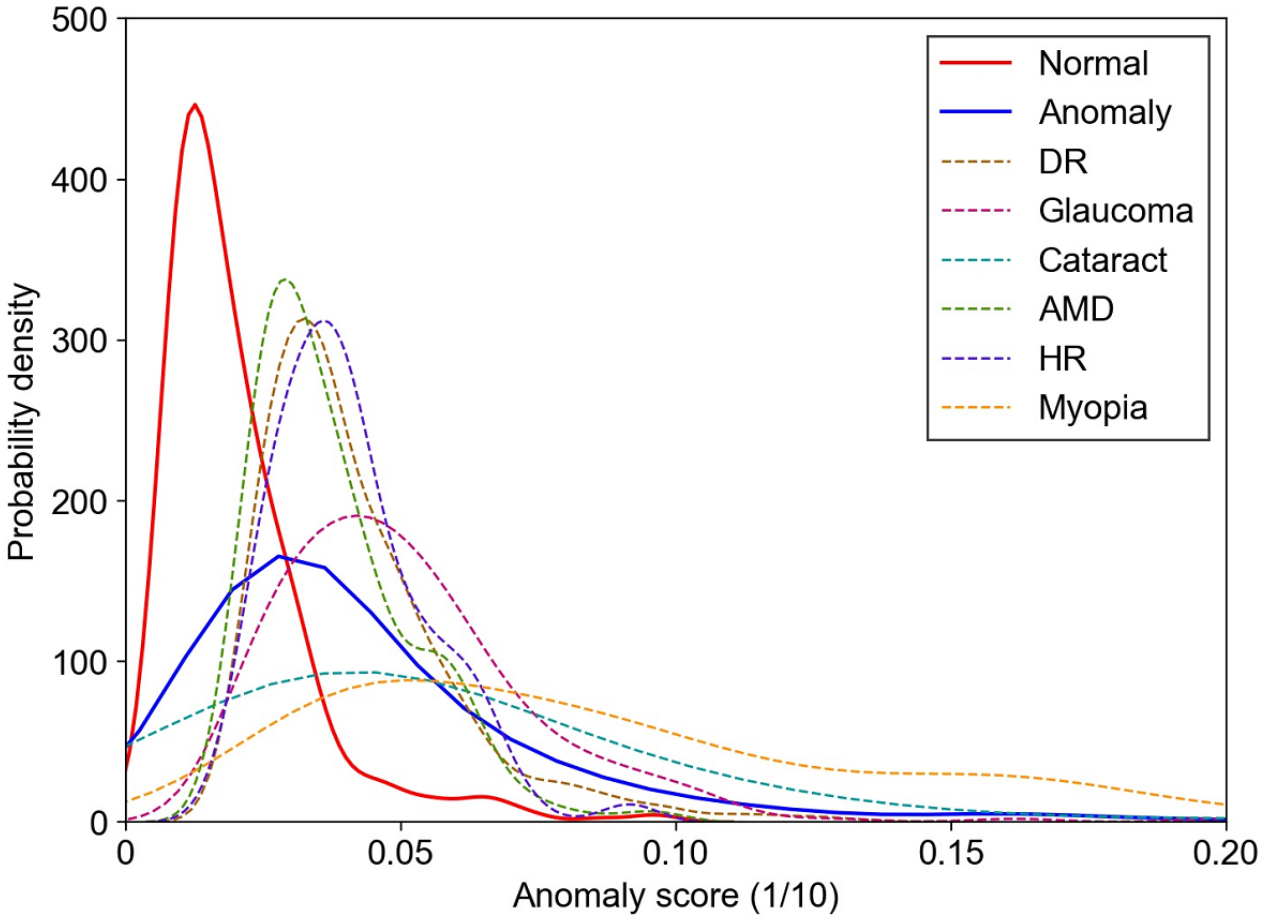
[b]The Wilson score was applied to calculate the Wilson 95% CI for accuracy, sensitivity, and specificity.

Figure 4. ROC and AUC of the anomaly detection model for detecting 6 ocular diseases. AMD: age-related macular degeneration, AUC: area under the receiver operating characteristic curve, DR: diabetic retinopathy, HR: hypertensive retinopathy, and ROC: receiver operating characteristic curve.



XSL•FO
RenderX

Figure 5. The probability density curves of the anomaly scores for the Ocular Disease Intelligent Recognition-5K test set. The solid red curve indicates the anomaly score distribution of normal fundus images, and the solid blue curve indicates the anomaly score distribution of all abnormal fundus images. The dotted line represents the anomaly score distribution of 6 types of abnormal fundus images. AMD: age-related macular degeneration, and DR: diabetic retinopathy, HR: hypertensive retinopathy.



Then, we assessed the performance of the model in detecting fundus images of referable DR and vision-threatening DR by using the Messidor-2 data set. Referable DR was defined as moderate DR or worse (severe DR and proliferative DR), and vision-threatening DR was defined as severe or proliferative DR. The AUC of the AD model was 0.908 for referable DR and 0.926 for vision-threatening DR, both of which were higher than the AUC of 0.868 for any DR (Table 4 and Figure 3).

Finally, we evaluated the model's performance in detecting 36 abnormal findings or diseases on retinal fundus images by using the JSIEC1000 data set. In detecting any abnormal findings or diseases, the model's AUC was 0.895 with an accuracy of 82.35%, sensitivity of 82.69%, and specificity of 82.99%. For each of the 36 abnormal findings or diseases, the AUC values ranged from 0.630 to 0.987, in which 23 of 36 categories had AUCs greater than 0.9, 9 categories had AUCs ranging from 0.8 to 0.9, and 4 categories had AUCs less than 0.8 (Multimedia Appendix 4).

Table 4. The Messidor-2 test set: area under the receiver operating characteristic curve, accuracy, sensitivity, and specificity of the anomaly detection model in detecting diabetic retinopathy of different severities.

Severity of diabetic retinopathy	Area under the receiver operating characteristic curve (95% CI)[a]	Proportion (%) (95% CI)[b]		
		Accuracy	Sensitivity	Specificity
Any diabetic retinopathy	0.868 (0.854-0.880)	79.45 (78.36-80.52)	79.47 (78.12-80.77)	79.41 (77.49-81.24)
Referable diabetic retinopathy	0.908 (0.899-0.916)	83.38 (81.49-84.84)	83.36 (81.77-84.84)	83.39 (81.87-84.81)
Vision-threatening diabetic retinopathy	0.926 (0.918-0.933)	86.40 (81.45-88.19)	86.38 (81.41-90.19)	86.43 (85.02-87.73)

[a]The binomial exact 95% CI was calculated for each are under the receiver operating characteristic curve.

[b]The Wilson score was applied to calculate the Wilson 95% CI for accuracy, sensitivity, and specificity.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, a GAN-based AD approach was developed to detect ocular diseases or abnormalities by using only normal fundus images. Four data sets containing over 53,236 color fundus images from various geographic and ethnic groups were applied for model training, validation, and internal testing, along with 4 data sets for external testing. Our results show that our approach achieved an AUC of 0.896, sensitivity of 80.69%, and specificity of 81.00% in detecting abnormal fundus images in the internal test set, and the developed model showed stable and consistent performance in the primary external test set, Local-2. This study further analyzed the model's ability to detect DR, glaucoma, cataracts, AMD, HR, and myopia and compared the performance of the model in detecting different severities of DR. To our knowledge, this is the first study using large-scale data sets of color fundus images to detect various ocular diseases by using an AD approach. The contribution of the AD model we established is to automatically detect abnormal fundus images so that the ophthalmologist can focus on the diagnosis based on abnormal fundus images and avoid spending extensive time and effort on masses of normal fundus images.

Comparison With Existing Studies

Previous studies have applied AD to screen eye disorders; however, most of those studies were based on OCT images. Retinal fundus photography is simpler to operate and more cost-effective than OCT, which renders it suitable for early screening of ocular disease [40]. Thus far, only Ouardini et al [34] have used AD with regard to color retinal fundus images; however, the 2 data sets they used were small, and only 1 type of fundus abnormality (retinopathy of prematurity or DR) was included in each data set. In our study, we developed an AD model based on 4 large-scale data sets derived from clinical or population screening and conducted external validation on 4 independent data sets, which ensured the robustness and generalizability of the model. Hence, this study substantially complements the findings of previous studies.

In recent years, there has been a surge in supervised deep learning studies for classifying fundus images, and many studies have achieved excellent performance. For example, Ting et al [41] trained a deep learning system to detect specific ocular diseases through binary classification tasks, achieving 90.5% sensitivity and 91.6% specificity for the detection of referable DR, 96.4% sensitivity and 87.2% specificity for possible glaucoma, and 93.2% sensitivity and 88.7% specificity for AMD, which is markedly higher than those of our method. The main reason may be that we used unsupervised learning; the model only learns the distribution pattern of normal images in model training, rather than the distributions of normal and abnormal images simultaneously as in supervised learning [42].

However, the unsupervised AD approach applied in our study has several specific advantages. First, model fitting does not require any labeled abnormal fundus images for training, which greatly reduces the cost of image annotation. Second, the AD model from our study can theoretically detect all classes of abnormal fundus images, including those of rare ocular diseases, while supervised learning is limited to detecting the types of abnormal images used in model training. Third, owing to the existence of various types of abnormal fundus images in clinical practice, the applicability of binary classification models is very low. For instance, the premise for a binary classification model to be applicable for distinguishing between normal and DR images is that the fundus image to be discriminated by the model must be either a normal fundus image or a DR fundus image. Any other type of image will inevitably lead to incorrect classification results. However, AD models do not present such issues.

Limitations

This study has some limitations of note. First, the model we established in this study has lower precision than that achievable by a supervised learning model. Nevertheless, the detection performance of the AD algorithm remained clinically acceptable and highly reproducible both in the internal test set and in external data sets. For DR screening, international guidelines recommended a minimum sensitivity of 60% (Australia) to 80% (United Kingdom) [43,44]. In clinical applications, to ensure that the model has a minimal false-negative rate, the sensitivity of the model can be increased by lowering the threshold of anomaly scores. Second, this model can only screen out abnormal fundus images and cannot directly provide specific diagnoses, which still require an ophthalmologist for completion. As such, the model is not capable of making a fully automated diagnosis, and its primary function is to allow ophthalmologists to focus more on the fundus images of possible lesions. The AD model is well suited for fundus screening in the general population, which predominately includes normal fundus images. Owing to the wide variety of existing ocular diseases, the use of fundus images alone is not a sufficient basis for accurate diagnosis of ocular diseases, and the model also requires information regarding the patient's medical history and clinical findings. Furthermore, the AD approach can detect anomalies only at the image level, not at the pixel level, and cannot show the specific locations of anomalies (eg, hard exudate, retinal vein occlusion, and macular hole) with heatmaps. The inherent black-box nature of deep learning may affect its acceptance for clinical use by ophthalmologists [45].

Conclusion

In conclusion, we developed and evaluated a cost-effective and time-efficient AD model to screen for ocular diseases or abnormalities, which showed high sensitivity and specificity. Further studies are required to determine the feasibility of applying this algorithm for clinical diagnosis or screening and to determine whether the use of this algorithm could lead to improved care and outcomes, compared to the current diagnostic workflow.

XSL•FO
RenderX

Acknowledgments

This study was supported by funds from the National Natural Science Foundation of China (award# 81530087, award# 81703318, and award# 82003559) and the Key Projects of Science and Technology Plan from the Beijing Municipal Education Commission (award# KZ201810025031).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pipeline of image pre-processing.
[DOCX File , 503 KB - jmir_v23i7e27822_app1.docx]

Multimedia Appendix 2
Comparison of area under the receiver operating characteristic curve (AUC) with varying input image resolution for AD model.
[DOCX File , 45 KB - jmir_v23i7e27822_app2.docx]

Multimedia Appendix 3
Modeling and inference scheme of the Skip-GANomaly algorithm.
[DOCX File , 20 KB - jmir_v23i7e27822_app3.docx]

Multimedia Appendix 4
The anomaly detection performance of Skip-GANomaly model for 36 categories abnormal fundus image using JISEC1000 dataset.
[DOCX File , 17 KB - jmir_v23i7e27822_app4.docx]

References

1. Blindness and vision impairment. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment [accessed 2021-06-24]
2. Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health 2017 Sep;5(9):e888-e897 [FREE Full text] [doi: 10.1016/S2214-109X(17)30293-0] [Medline: 28779882]
3. Diabetic Retinal Screening, Grading, Monitoring and Referral Guidance. New Zealand Government: Ministry of Health. 2016. URL: https://www.health.govt.nz/publication/diabetic-retinal-screening-grading-monitoring-and-referral-guidance [accessed 2021-06-24]
4. Panwar N, Huang P, Lee J, Keane PA, Chuan TS, Richhariya A, et al. Fundus Photography in the 21st Century--A Review of Recent Technological Advances and Their Implications for Worldwide Healthcare. Telemed J E Health 2016 Mar;22(3):198-208 [FREE Full text] [doi: 10.1089/tmj.2015.0068] [Medline: 26308281]
5. Muqit MMK, Kourgialis N, Jackson-deGraffenried M, Talukder Z, Khetran ER, Rahman A, et al. Trends in Diabetic Retinopathy, Visual Acuity, and Treatment Outcomes for Patients Living With Diabetes in a Fundus Photograph-Based Diabetic Retinopathy Screening Program in Bangladesh. JAMA Netw Open 2019 Nov 01;2(11):e1916285 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.16285] [Medline: 31774523]
6. Massin P, Marre M. Fundus photography for the screening for diabetic retinopathy. Diabetes Metab 2002 Apr;28(2):151-155. [doi: 10.2337/diacare.25.4.804] [Medline: 11976569]
7. Seddon JM, Sharma S, Adelman RA. Evaluation of the clinical age-related maculopathy staging system. Ophthalmology 2006 Feb;113(2):260-266. [doi: 10.1016/j.ophtha.2005.11.001] [Medline: 16458093]
8. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep 2018 Oct 02;8(1):14665 [FREE Full text] [doi: 10.1038/s41598-018-33013-w] [Medline: 30279554]
9. Ferraro JG, Pollard T, Muller A, Lamoureux EL, Taylor HR. Detecting cataract causing visual impairment using a nonmydriatic fundus camera. Am J Ophthalmol 2005 Apr;139(4):725-726. [doi: 10.1016/j.ajo.2004.09.083] [Medline: 15808179]
10. Phan TV, Seoud L, Chakor H, Cheriet F. Automatic Screening and Grading of Age-Related Macular Degeneration from Texture Analysis of Fundus Images. J Ophthalmol 2016;2016:5893601 [FREE Full text] [doi: 10.1155/2016/5893601] [Medline: 27190636]
11. Zhang H, Niu K, Xiong Y, Yang W, He Z, Song H. Automatic cataract grading methods based on deep learning. Comput Methods Programs Biomed 2019 Dec;182:104978. [doi: 10.1016/j.cmpb.2019.07.006] [Medline: 31450174]

XSL•FO
RenderX

12. Lovis C. Unlocking the Power of Artificial Intelligence and Big Data in Medicine. J Med Internet Res 2019 Nov 08;21(11):e16607 [FREE Full text] [doi: 10.2196/16607] [Medline: 31702565]
13. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial Intelligence in Clinical Health Care Applications: Viewpoint. Interact J Med Res 2019 Apr 05;8(2):e12100 [FREE Full text] [doi: 10.2196/12100] [Medline: 30950806]
14. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
15. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
16. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. Ophthalmology 2017 Jul;124(7):962-969. [doi: 10.1016/j.ophtha.2017.02.008] [Medline: 28359545]
17. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. JAMA Ophthalmol 2017 Nov 01;135(11):1170-1176 [FREE Full text] [doi: 10.1001/jamaophthalmol.2017.3782] [Medline: 28973096]
18. Pead E, Megaw R, Cameron J, Fleming A, Dhillon B, Trucco E, et al. Automated detection of age-related macular degeneration in color fundus photography: a systematic review. Surv Ophthalmol 2019;64(4):498-511 [FREE Full text] [doi: 10.1016/j.survophthal.2019.02.003] [Medline: 30772363]
19. Fu H, Cheng J, Xu Y, Zhang C, Wong DWK, Liu J, et al. Disc-Aware Ensemble Network for Glaucoma Screening From Fundus Image. IEEE Trans Med Imaging 2018 Nov;37(11):2493-2501. [doi: 10.1109/TMI.2018.2837012] [Medline: 29994764]
20. Liu J, Wong D, Lim J, Tan N, Zhang Z, Li H, et al. Detection of Pathological Myopia by PAMELA with Texture-Based Features through an SVM Approach. J Healthc Eng 2010 Jan;1(1):1-12. [doi: 10.1260/2040-2295.1.1.1]
21. Akçay S, Atapour-Abarghouei A, Breckon T. Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. 2019 Presented at: 2019 International Joint Conference on Neural Networks; 2019; Budapest. [doi: 10.1109/ijcnn.2019.8851808]
22. Markou M, Singh S. Novelty detection: a review—part 1: statistical approaches. Signal Processing 2003 Dec;83(12):2481-2497. [doi: 10.1016/j.sigpro.2003.07.018]
23. Markou M, Singh S. Novelty detection: a review—part 2: neural network based approaches. Signal Processing 2003 Mar;2003(3):2499-2521. [doi: 10.1016/S1353-4858(03)00313-1]
24. Hodge V, Austin J. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review 2004 Oct;22(2):85-126. [doi: 10.1023/B:AIRE.0000045502.10941.a9]
25. Pimentel MA, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Signal Processing 2014 Jun;99:215-249. [doi: 10.1016/j.sigpro.2013.12.026]
26. Abdallah A, Maarof MA, Zainal A. Fraud detection system: A survey. J Netw Comput Appl 2016 Jun;68:90-113. [doi: 10.1016/j.jnca.2016.04.007]
27. Ahmed M, Naser Mahmood A, Hu J. A survey of network anomaly detection techniques. J Netw Comput Appl 2016 Jan;60:19-31. [doi: 10.1016/j.jnca.2015.11.016]
28. Ahmed M, Mahmood AN, Islam MR. A survey of anomaly detection techniques in financial domain. Future Gener Comput Syst 2016 Feb;55:278-288. [doi: 10.1016/j.future.2015.01.001]
29. Kiran B, Thomas D, Parakkal R. An Overview of Deep Learning Based Methods for Unsupervised and Semi-Supervised Anomaly Detection in Videos. J Imaging 2018 Feb 07;4(2):36. [doi: 10.3390/jimaging4020036]
30. Seebock P, Waldstein SM, Klimscha S, Bogunovic H, Schlegl T, Gerendas BS, et al. Unsupervised Identification of Disease Marker Candidates in Retinal OCT Imaging Data. IEEE Trans Med Imaging 2019 Apr;38(4):1037-1047. [doi: 10.1109/TMI.2018.2877080] [Medline: 30346281]
31. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. 2017 Presented at: 25th International Conference on Information Processing in Medical Imaging; June 25-30, 2017; Boone, NC.
32. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med Image Anal 2019 May;54:30-44. [doi: 10.1016/j.media.2019.01.010] [Medline: 30831356]
33. Zhou K, Gao S, Cheng J, Gu Z, Fu H, Tu Z, et al. Sparse-Gan: Sparsity-Constrained Generative Adversarial Network for Anomaly Detection in Retinal OCT Image. 2020 Presented at: IEEE 17th International Symposium on Biomedical Imaging; April 04-07,2020; Iowa City. [doi: 10.1109/isbi45749.2020.9098374]
34. Ouardini K, Yang H, Unnikrishnan B, Romain M, Garcin C, Zenati H, et al. Towards Practical Unsupervised Anomaly Detection on Retinal Images. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data. Cham: Springer International Publishing; 2019:225-234.

XSL•FO
RenderX

35. Abràmoff MD, Folk JC, Han DP, Walker JD, Williams DF, Russell SR, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013 Mar;131(3):351-357. [doi: 10.1001/jamaophthalmol.2013.1743] [Medline: 23494039]
36. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, et al. Feedback on a publicly distributed image database: the Messidor database. Image Anal Stereol 2014 Aug 26;33(3):231. [doi: 10.5566/ias.1155]
37. Wilkinson CP, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 2003 Sep;110(9):1677-1682. [doi: 10.1016/S0161-6420(03)00475-5] [Medline: 13129861]
38. Davletshina D, Melnychuk V, Tran V, Singla H, Berrendorf M, Faerman E, et al. Unsupervised Anomaly Detection for X-Ray Images. ArXiv 2020;abs/2001.10883:1-22 (forthcoming) [FREE Full text]
39. Youden WJ. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
40. Hagiwara Y, Koh JEW, Tan JH, Bhandary SV, Laude A, Ciaccio EJ, et al. Computer-aided diagnosis of glaucoma using fundus images: A review. Comput Methods Programs Biomed 2018 Oct;165:1-12. [doi: 10.1016/j.cmpb.2018.07.012] [Medline: 30337064]
41. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 2017 Dec 12;318(22):2211-2223 [FREE Full text] [doi: 10.1001/jama.2017.18152] [Medline: 29234807]
42. Sathya R, Abraham A. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. IJARAI 2013;2(2):34-38. [doi: 10.14569/IJARAI.2013.020206]
43. Chakrabarti R, Harper CA, Keeffe JE. Diabetic retinopathy management guidelines. Expert Rev Ophthalmol 2014 Jan 09;7(5):417-439. [doi: 10.1586/eop.12.52]
44. Diabetic eye screening: commission and provide. National Health Service (NHS) Diabetic Eye Screening Programme and Population Screening Programmes. 2015. URL: https://www.gov.uk/government/collections/diabetic-eye-screening-commission-and-provide [accessed 2021-06-24]
45. Wong TY, Bressler NM. Artificial Intelligence With Deep Learning Technology Looks Into Diabetic Retinopathy Screening. JAMA 2016 Dec 13;316(22):2366-2367. [doi: 10.1001/jama.2016.17563] [Medline: 27898977]

Abbreviations

AD: anomaly detection
AI: artificial intelligence
AMD: age-related macular degeneration
AUC: area under the receiver operating characteristic curve
DR: diabetic retinopathy
GAN: generative adversarial network
HR: hypertensive retinopathy
OCT: optical coherence tomography
ODIR: Ocular Disease Intelligent Recognition

Edited by R Kukafka; submitted 08.02.21; peer-reviewed by J Huang, W Yang; comments to author 23.03.21; revised version received 07.05.21; accepted 24.05.21; published 13.07.21.

Please cite as:
Han Y, Li W, Liu M, Wu Z, Zhang F, Liu X, Tao L, Li X, Guo X
Application of an Anomaly Detection Model to Screen for Ocular Diseases Using Color Retinal Fundus Images: Design and Evaluation Study
J Med Internet Res 2021;23(7):e27822
URL: https://www.jmir.org/2021/7/e27822
doi:10.2196/27822
PMID:34255681

©Yong Han, Weiming Li, Mengmeng Liu, Zhiyuan Wu, Feng Zhang, Xiangtong Liu, Lixin Tao, Xia Li, Xiuhua Guo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

XSL•FO
RenderX

Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Relative Performance of Machine Learning and Linear Regression in Predicting Quality of Life and Academic Performance of School Children in Norway: Data Analysis of a Quasi-Experimental Study

Robert Froud[1,2], BSc, MSc, PhD; Solveig Hakestad Hansen[1], BSc; Hans Kristian Ruud[1], MSc; Jonathan Foss[3], PhD; Leila Ferguson[1], PhD; Per Morten Fredriksen[1], PhD

[1]School of Health Sciences, Kristiania University College, Oslo, Norway

[2]Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[3]Department of Computer Science, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Robert Froud, BSc, MSc, PhD
School of Health Sciences
Kristiania University College
Prinsens Gate 7-9
Oslo, 0107
Norway
Phone: 47 1732494636
Email: rob.froud@kristiania.no

Abstract

Background: Machine learning techniques are increasingly being applied in health research. It is not clear how useful these approaches are for modeling continuous outcomes. Child quality of life is associated with parental socioeconomic status and physical activity and may be associated with aerobic fitness and strength. It is unclear whether diet or academic performance is associated with quality of life.

Objective: The purpose of this study was to compare the predictive performance of machine learning techniques with that of linear regression in examining the extent to which continuous outcomes (physical activity, aerobic fitness, muscular strength, diet, and parental education) are predictive of academic performance and quality of life and whether academic performance and quality of life are associated.

Methods: We modeled data from children attending 9 schools in a quasi-experimental study. We split data randomly into training and validation sets. Curvilinear, nonlinear, and heteroscedastic variables were simulated to examine the performance of machine learning techniques compared to that of linear models, with and without imputation.

Results: We included data for 1711 children. Regression models explained 24% of academic performance variance in the real complete-case validation set, and up to 15% in quality of life. While machine learning techniques explained high proportions of variance in training sets, in validation, machine learning techniques explained approximately 0% of academic performance and 3% to 8% of quality of life. With imputation, machine learning techniques improved to 15% for academic performance. Machine learning outperformed regression for simulated nonlinear and heteroscedastic variables. The best predictors of academic performance in adjusted models were the child's mother having a master-level education (P<.001; β=1.98, 95% CI 0.25 to 3.71), increased television and computer use (P=.03; β=1.19, 95% CI 0.25 to 3.71), and dichotomized self-reported exercise (P=.001; β=2.47, 95% CI 1.08 to 3.87). For quality of life, self-reported exercise (P<.001; β=1.09, 95% CI 0.53 to 1.66) and increased television and computer use (P=.002; β=−0.95, 95% CI −1.55 to −0.36) were the best predictors. Adjusted academic performance was associated with quality of life (P=.02; β=0.12, 95% CI 0.02 to 0.22).

Conclusions: Linear regression was less prone to overfitting and outperformed commonly used machine learning techniques. Imputation improved the performance of machine learning, but not sufficiently to outperform regression. Machine learning techniques outperformed linear regression for modeling nonlinear and heteroscedastic relationships and may be of use in such cases. Regression with splines performed almost as well in nonlinear modeling. Lifestyle variables, including physical exercise, television and computer use, and parental education are predictive of academic performance or quality of life. Academic performance is associated with quality of life after adjusting for lifestyle variables and may offer another promising intervention target to improve quality of life in children.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e22021) doi:10.2196/22021

KEYWORDS

modelling; linear regression; machine learning; artificial intelligence; quality of life; academic performance; continuous/quasi-continuous health outcomes

Introduction

In trials and quasi-experimental designs, reported sample sizes range from less than 100 to several thousand [1]. Linear regression approaches are widely used for modeling continuous outcome data in such studies [2]. Processor advancements, data abundance, and routine data collection have cultivated a general rise in popularity of artificial intelligence or machine learning techniques. In contrast to regression, the use of machine learning techniques requires making fewer assumptions about data structure [3]. Machine learning techniques have been used extensively in areas such as biomedicine and, to a lesser extent, in areas such as chronic disease, pain, psychology, and sociology, where data have not typically been available in such abundance [4-6]. Machine learning techniques have yielded useful health classification models [7,8]. Numerous comparisons exist between machine learning techniques and traditional logistic or multinomial logit regression, demonstrating that approaches can yield similar performance and highlighting a risk of overfitting in machine learning techniques [9]. However, few comparisons exist between machine learning techniques and linear regression for continuous outcomes in health data sets, and where such comparisons have been made, sample sizes have been small [10,11].

Quality of life is an important health outcome in trials [2,12]. Child quality of life is associated with parental socioeconomic status and activity levels [13-16]. Diet is associated with child mental health, but the nature of the relationship between diet and child quality of life is less clear [17,18]. It has been suggested that aerobic fitness and muscular strength are positively associated with child quality of life [13,19]. The extent to which academic performance and quality of life are associated is also unclear. Known predictors of academic performance include parental socioeconomic status, child IQ, and activity levels, and there is some evidence of association with diet [20-22]. Thus, any relationship between quality of life and academic performance may be confounded by common associations with socioeconomic status, activity, and diet. Our aims were to examine the performance of linear regression and common machine learning techniques; the extent to which lifestyle variables (including physical activity, aerobic fitness, muscular strength, diet) and parental education are predictive of academic performance and quality of life; and the association between academic performance and quality of life, after adjusting for confounding variables, using a relatively large data set with continuous health outcomes.

Methods

Data Set

We used data from fifth-year students attending 9 schools in Norway between 2015 and 2019, within the Health Oriented Pedagogical Project (HOPP), which is an ongoing quasi-experimental study (ClinicalTrials.gov; NCT02495714) in which data up to 2019 were captured [23]. Schools were allocated to receive intervention (n=7) or usual curriculum (n=2). In intervention schools, child activity was increased by 225 minutes per week and an activity-based learning component (emphasizing mathematics and language studies, including English) was undertaken during the physical activity [23]. Both parent and child quality of life was measured using the Norwegian version of the Inventory of Life Quality [24]. The Norwegian Inventory of Life Quality has good internal consistency (normative 11- to 12-year-old children: Cronbach α=.82; parents: Cronbach α=.80) and good test–retest reliability in Norwegian children (normative 11- to 14-year-old children: intraclass correlation coefficient 0.86) [25]. In parents, test–retest has been reported as satisfactory, although we found no reports of a published intraclass correlation coefficient [25]. The Inventory of Life Quality spans domains of perceived school performance, family relations, peer relations, autonomy in play, physical health, mental health, and global assessment of well-being and uses a measurement scale of 0 to 100, where higher scores indicate greater quality of life. Academic performance was measured using the Norwegian Directorate for Education and Training's compulsory National Academic Tests for fifth year students. We had access to reading, mathematics, and English test results; each academic subject was measured on a quasi-continuous scale (ranging from 0 to 100). Because we were interested in general academic performance, we used the average of these tests [26].

Physical activity level (defined by movement counts per minute: sedentary 0-99, light 100-1999, moderate 2000-4999, and hard or vigorous ≥5000 [13], while a monitor was worn between 8 hours and 6 days), percentage of time spent at each activity level, and average moderate-to-vigorous physical activity level (the sum of the minutes spent in moderate-to-vigorous activity divided by the number of valid monitored days) [27]; weight; height; blood pressure; waist circumference; muscle mass; percentage body fat; hand strength; aerobic fitness (Andersen intermittent running test [28]); executive functions (Stroop test [29,30]); parental education (university education or not; masters level or above or not); and lifestyle (self-reported diet, physical activity, and health questions from the Ungkost-2000 questionnaire [31]) data were included as predictor variables. Where there were missing observations in year 5 Ungkost variables, we carried forward observations from the same pupils in year 4.

Modeling Approaches

We split the data set randomly into training (70%) and validation (30%) sets in order to train models and subsequently evaluate performance. We expected missing data (approximately 20% overall, with few variables >50%). Full imputation may often be performed with machine learning techniques regardless of

XSL•FO
RenderX

the extent of missing data or whether or not data are missing at random. We performed a sensitivity analysis using single-mean imputation for continuous predictor variables and mode for nonbinary or categorical predictors (stratified by school) under the assumption that observations were missing at random. We tested this assumption for variables in final models by fitting a dummy variable for variable missingness, examining effect on outcome using 2-tailed independent *t* tests. In addition, we simulated variables with no missing data. We first examined strengths and limitations of different approaches, modeling academic performance with worked examples, and then modeled child quality of life.

Regression Modeling

We took a pragmatic approach to regression modeling that we judged to approximate best practice. In cases of high between-predictor correlations (ρ>0.75), we selected 1 variable for modeling. In the absence of strong clinical or theoretical indications, we chose the variable that explained the most variance. To enable comparisons to regression approaches in which individuals are clustered by site, we fitted linear mixed models with a random intercept by school. We also built nonhierarchical models, without this random effect, to compare adjusted R^2 like-for-like with machine learning techniques (in which clustering was not nominated). To facilitate comparison of residual mean square error (RMSE), we standardized variables by subtracting the mean and dividing by the standard deviation, which is required by machine learning techniques. For curvilinear relationships, we explored fitting polynomial terms. In the case of truly nonlinear relationships—variables that are not well modeled with a single linear predictor (notwithstanding polynomial terms)—we fitted splines (ie, piecewise fitting of models) [32].

The diet and lifestyle variables from the Ungkost-2000 questionnaire have multiple quasi-continuous responses (eg, for sugared soda consumption, response options ranged from 'Never/rarely' and increased incrementally over 7-levels to a maximum that indicated >7 glasses per day). Where responses were normally distributed, we treated the variables as quasi-continuous. If distributions did not satisfy normality criteria, we dichotomized variables using a cut-point [33]. Variables with significant crude effects were considered for an adjusted model. We took a manual approach to model building, using a combination of the lowest Akaike Information Criterion and variables that we judged to be clinically or theoretically useful for outcome prediction [34]. When modeling academic performance, in order to facilitate performance comparisons with partly automated machine learning techniques, we did not favor modifiable exposures, but instead, favored those we judged would explain the most variance. For quality of life, we built 2 models: (1) optimized for prediction and (2) based on modifiable exposures. Models for sensitivity analyses with imputed data were built independently.

Machine Learning Techniques

We evaluated the performance of 4 machine learning techniques (Table 1) [35,36]. We selected machine learning techniques that were able to be used with continuous outcome measures (and not only binary or categorical), appear commonly in health research literature, and we judged health researchers would find comparisons useful. It is beyond the scope of this paper to explain each technique in detail; however, overviews are provided in Table 1.

Variables that it did not make sense to include were removed (eg, age, since participants were from the same school year). We set each approach to start with a null model and successively added variables that provided the best improvement, measured by RMSE in cross-validation [35]. We only selected tuning characteristics, such as the optimum value of *k* in k-nearest neighbor models, or optimum decay and threshold activation levels in neural network models, after graphical assessment. For machine learning techniques, we did not dichotomize nonnormal diet and lifestyle variables, since machine learning techniques are not sensitive to normality.

Table 1. Machine learning techniques that were evaluated in this study.

Algorithm	Description
k-Nearest neighbors	A classification technique that assigns class or predicts a continuous value based on the classes or values of *k* nearest neighbors.
Neural network	A technique in which artificial neuron cores are connected with *n* input channels, inputs are weighted and summed, and the output (if above an activation threshold) feeds into another neuron in a deeper hidden layer. This deeper neuron receives multiple inputs from each neuron in the layer above, and communicates output with either another hidden layer, or an output layer. Synaptic weights in this structure are determined by back propagation, based on error, until convergence is reached.
Random forest	An iteratively grown set of decision trees, where each tree outputs outcome means, with branches split by variable characteristics, and where each tree is formed from randomly bootstrapping data, with averages taken from all trees.
Support vector machine	A technique that minimizes error to individualize a hyperplane.

Simulations

We simulated data to explore types of relationship that were not present within our real data, but which we reasoned, may perform better with either regression or machine learning techniques. We simulated, without missing data, (1) a variable with a quadratic relationship with academic performance; (2) a variable with a true nonlinear relationship with academic performance; and (3) a variable with marked heteroscedasticity (ie, changing variance) with respect to academic performance (we acknowledge this is a technical violation of regression; therefore, we recorded R^2 and RMSE rather than standard error

XSL•FO
RenderX

terms). We permitted slight heteroscedasticities to remain in the first 2 simulations to approximate limits of real-life pragmatic decisions. We expected curvilinear simulation to favor regression, since we reasoned it would be modeled well with polynomial terms; nonlinear simulation to favor machine learning techniques, or linear regression with splines, since truly nonlinear relationships are not conducive to modeling by a single linear predictor; and heteroscedastic simulation to favor machine learning techniques, since modeling is not derived using minimum squared error, which in the presence of heteroscedasticity would no longer be the best estimator.

Performance Comparisons and Using Worked Examples for Modeling Quality of Life

To compare performance, we calculated RMSE and R^2 using predicted observations from training sets and observed observations from validation sets (Multimedia Appendix 1). Informed by findings from modeling academic performance, we judged the most appropriate modeling technique for quality of life, and to confirm that we had made the correct choice, we compared the performance of the approaches that we selected with those that we did not select.

To aid interpretation of adjusted regression model outputs for those unfamiliar with the outcome scales, we calculated Cohen *d* for our judgements of clinically intuitive predictor magnitudes, by outcome variable; where *d* may be interpreted by thresholds of small (0.2), medium (0.5), and large (0.8) effects [37].

All analyses were performed using Stata (version 15.1; StataCorp LLC) and R (version 3.6; R Foundation for Statistical Computing). The HOPP project received approval from the Norwegian Regional Ethical Committee (2014/2064/REK south-east), and parents of all children provided written informed consent for their child's participation.

Results

Overview

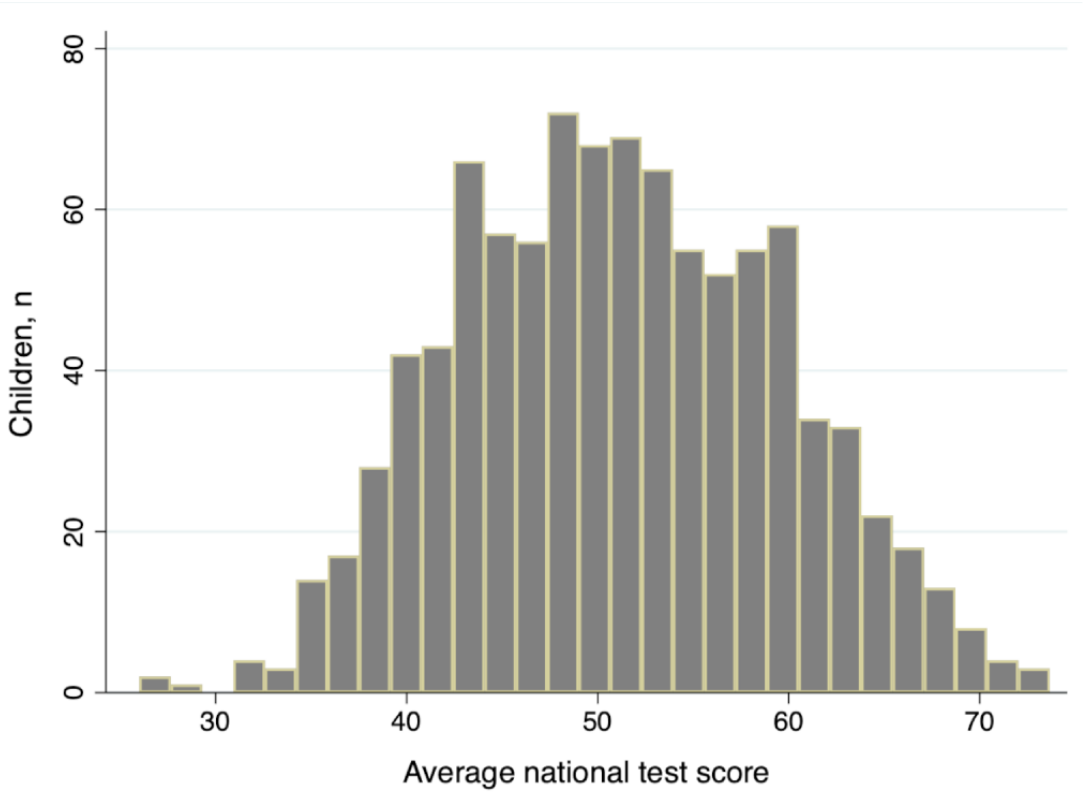
Data comprised outcomes from 1711 year 5 (11- and 12-year-old) children (Tables S1 and S2 in Multimedia Appendix 1), of whom 1368 (80.0%) had completed National Test outcomes and 1560 (91.6%) had completed quality of life outcomes. Missing data ranged from 4% to 81%, by variable. Our training and validation data sets had data from 1205 and 506 children, respectively.

Academic Performance and Simulated Data

Academic performance was approximately normally distributed (Figure 1). From crudely modeled academic performance variables (Table S3 in Multimedia Appendix 1), we selected 7 variables for modeling (Table 2). We noted that after adjustment, dietary variables either explained too little variance or had too few observations for us to select for inclusion. Machine learning techniques did retain some dietary variables (Table S4 in Multimedia Appendix 1).

In real complete-case data, nonhierarchical and mixed models explained approximately 30% of the variance in the training set and 22% to 24% of the variance in the validation set (Table 3). Model residuals were normally distributed. Machine learning models explained between 13% and 63% of the variance in the training set and approximately 0% of the variance in validation (Table 3).

Figure 1. Histogram of average national test scores.



XSL•FO
RenderX

Table 2. Adjusted effects in selected mixed regression model for predicting academic performance.

Variable	β (95% CI)	n	*P* value
Stroop test congruent (milliseconds)	−0.0037 (−0.0047 to −0.0027)	384	<.001
Effect of master-level education for father	1.59 (−0.06 to 3.25)	384	.06
Effect of master-level education for mother	1.98 (0.25 to 3.71)	384	<.001
Average hand strength (kilograms)	0.21 (0.08 to 0.34)	384	.001
Hours of physical activity (self-reported; dichotomized)	2.47 (1.08 to 3.87)	384	.001
Effect of mother having higher education	1.82 (0.07 to 3.57)	384	.04
Hours of television per week (self-reported; 7-level quasi-continuous)	1.19 (0.25 to 3.71)	384	.03

Table 3. Performance indicators in real data and real data augmented with simulated data (quadratic, nonlinear, or heteroscedastic) for academic performance.

Model	Training (n=962)			Validation (n=406)		
	RMSE[a]	R^2 value[b]	n	RMSE	R^2 value[b]	n
Nonhierarchical linear model	0.81	0.30	384	0.85	0.22	163
Quadratic	0.45	0.78	384	0.40	0.83	163
Nonlinear	0.55	0.68	384	0.53	0.70	163
Heteroscedastic	0.53	0.70	384	0.61	0.61	163
Mixed model	0.83	0.30	384	0.86	0.24	163
Quadratic	0.46	0.79	384	0.39	0.84	163
Nonlinear	0.56	0.68	384	0.53	0.72	163
Heteroscedastic	0.54	0.70	384	0.62	0.62	163
Regression with splines	—[c]	—	—	—	—	—
Nonlinear	0.41	0.82	384	0.39	0.84	163
Random forest	0.61	0.62	121	0.95	−0.02	63
Quadratic	0.32	0.91	121	0.51	0.75	63
Nonlinear	0.36	0.89	121	0.57	0.64	63
Heteroscedastic	0.34	0.89	121	0.67	0.53	63
Support vector machine	0.55	0.63	116	0.89	−0.05	58
Quadratic	0.33	0.87	116	0.53	0.62	58
Nonlinear	0.46	0.77	116	0.77	0.18	58
Heteroscedastic	0.35	0.85	116	0.62	0.52	58
k-Nearest neighbors	0.90	0.13	133	1.02	−0.01	66
Quadratic	0.37	0.84	133	0.48	0.75	66
Nonlinear	0.41	0.81	133	0.48	0.75	66
Heteroscedastic	0.43	0.79	133	0.61	0.60	66
Neural network	0.73	0.35	124	1.03	−0.02	66
Quadratic	0.38	0.82	124	0.40	0.85	66
Nonlinear	0.41	0.79	124	0.46	0.79	66
Heteroscedastic	0.43	0.77	124	0.70	0.53	66

[a]RMSE: residual mean square error.

[b]Unlike unadjusted R^2, it is possible for adjusted R^2 values to be negative.

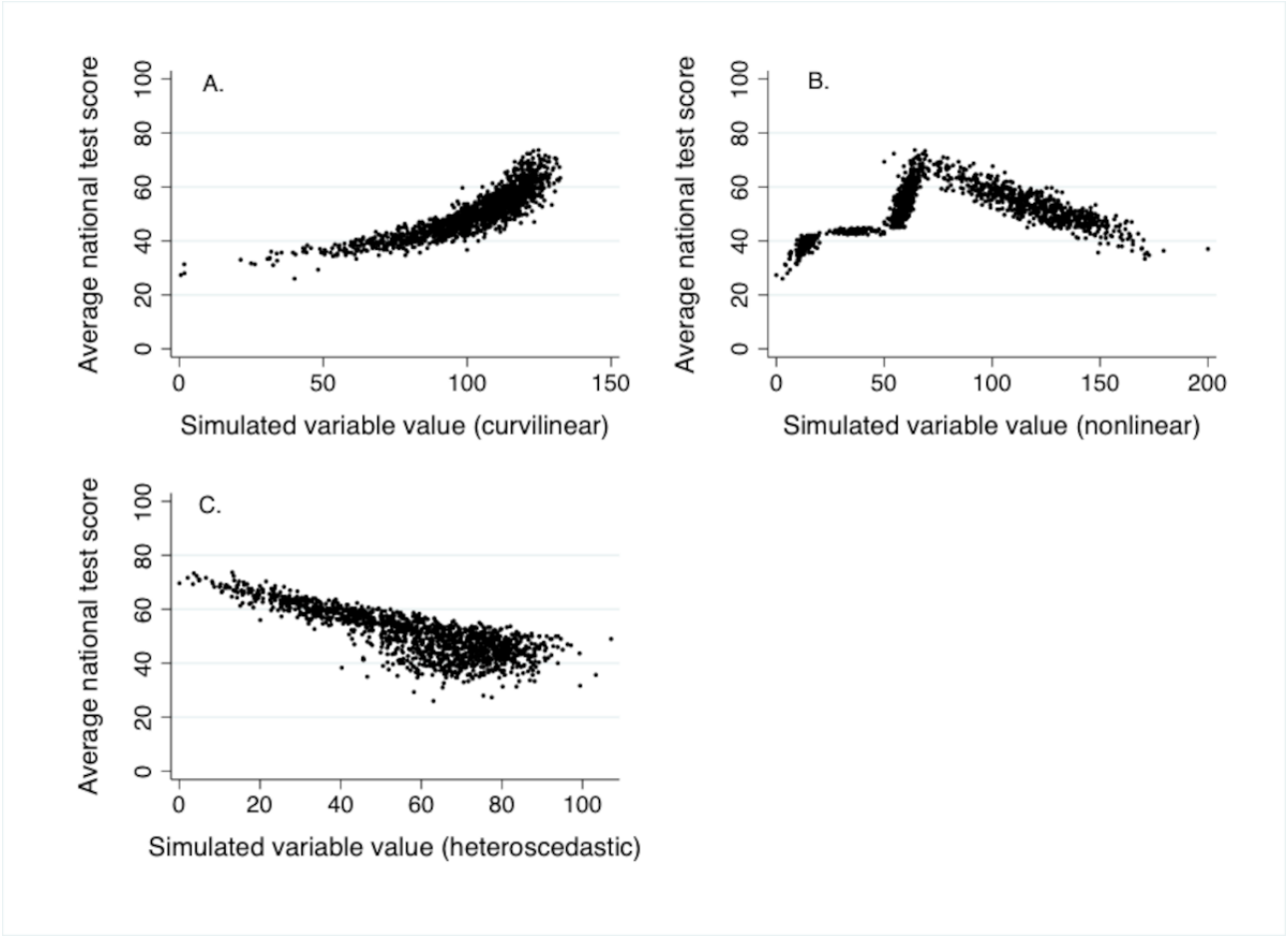
[c]Not performed.

XSL•FO
RenderX

Figure 2 shows scatter plots of academic performance and simulated variables. All had strong effects in regression models when modeled as quadratic, quadratic, and linear. Adding a simulated quadratic variable to crude regression models explained approximately 79% of the variance in the training set and 82% to 83% of the variance in the validation (Table 4). Corresponding machine learning models explained 80% to 94% of the variance in the training set and 78% to 83% of the variance in the validation set, with support vector machine and neural network performing best. The nonlinear simulation was the only one with a variable that had a nonlinear relationship with academic performance, and we fitted 4 splines. Regression with splines explained 83% of the variance in the training set and 85% of the variance in the validation set. Corresponding machine learning models explained 81% to 94% of variance in the training set and 81% to 86% of the variance in the validation set, with neural network performing best. Adding a simulated heteroscedastic variable to crude regression models explained 64% of variance in the training set and 62% of the variance in the validation set. Corresponding machine learning models explained 68% to 90% of the variance in the training set and 58% to 66% of the variance in the validation set, with neural network and support vector machine performing best.

Regression performed best for modeling real data augmented with simulations (Table 3). Regression with splines performed best when adding the nonlinear simulated variable. Table 5 shows machine learning performance improved after imputation; however, regression models outperformed machine learning. Regression models built using imputed data included 13 variables (Multimedia Appendix 1). Variables selected by machine learning techniques are shown in Table S4 in Multimedia Appendix 1. The missing at random assumption was widely acceptable, with 3 out of 35 variables selected for modeling (master's education or above for mother, master's education or above for father, and parent quality of life score) having an effect on academic performance.

Figure 2. Scatter plots of average national test score and simulated (A) curvilinear, (B) nonlinear, and (C) heteroscedastic variables.



XSL•FO
RenderX

Table 4. Crude performance of simulated variables.

Model	Training (n=962)			Validation (n=406)		
	RMSE[a]	R^2 value	n	RMSE[a]	R^2 value	n
Nonhierarchical linear model						
Quadratic	0.45	0.79	962	0.43	0.82	406
Nonlinear	0.56	0.68	962	0.58	0.68	406
Heteroscedastic	0.59	0.64	962	0.63	0.62	406
Mixed model						
Quadratic	0.45	0.79	962	0.43	0.83	406
Nonlinear	0.57	0.68	962	0.58	0.68	406
Heteroscedastic	0.59	0.64	962	0.63	0.62	406
Regression with splines						
Nonlinear	0.41	0.83	962	0.39	0.85	406
Random forest						
Quadratic	0.25	0.94	962	0.49	0.78	406
Nonlinear	0.24	0.94	962	0.45	0.81	406
Heteroscedastic	0.32	0.90	962	0.66	0.58	406
Support vector machine						
Quadratic	0.44	0.80	962	0.42	0.83	406
Nonlinear	0.44	0.81	962	0.43	0.82	406
Heteroscedastic	0.57	0.68	962	0.61	0.66	406
k-Nearest neighbors						
Quadratic	0.40	0.84	962	0.45	0.81	406
Nonlinear	0.34	0.88	962	0.43	0.82	406
Heteroscedastic	0.49	0.76	962	0.65	0.60	406
Neural network						
Quadratic	0.44	0.80	962	0.42	0.83	406
Nonlinear	0.40	0.84	962	0.38	0.86	406
Heteroscedastic	0.56	0.68	962	0.59	0.66	406

[a]RMSE: residual mean square error.

Table 5. Performance indicators for academic performance in sensitivity analyses (single-mean imputation).

Model	Training (n=962)			Validation (n=406)		
	RMSE[a]	R^2 value	n	RMSE[a]	R^2 value	n
Nonhierarchical linear model	0.88	0.20	962	0.92	0.15	406
Mixed model	0.89	0.21	962	0.92	0.18	406
Random forest	0.76	0.48	962	0.94	0.14	406
Support vector machine	0.82	0.32	962	0.95	0.12	406
k-Nearest neighbors	0.89	0.20	962	0.86	0.12	406
Neural network	0.90	0.18	962	0.97	0.09	406

[a]RMSE: residual mean square error.

XSL•FO
RenderX

Quality of Life

Despite a ceiling effect, we judged the distribution of child-reported quality of life (Figure 3) to be within limits of tolerance for untransformed parametric modeling (and we confirmed there was a normal distribution of residuals postmodeling). Since visual inspection revealed no nonlinear relationships, and only very slight heteroscedasticity at times, we judged regression modeling would perform best. We dichotomized 1 diet variable (fish oil consumption) based on crude effects (Table S5 in Multimedia Appendix 1). We selected a parsimonious 3-variable model (Regression model 1) on the basis of raw performance (Table S6 in Multimedia Appendix 1) and a second 4-predictor model (Regression model 2) using only variables with a high number of observations and representing modifiable risk factors (Table 6). When added, academic performance had a significant association with quality of life (P=.02), with an adjusted effect of 0.12 (95% CI 0.02 to 0.22). We did not include academic performance in our comparative model because it reduced observations and led to lower training R^2 values. Two of the machine learning techniques retained academic performance and several diet variables in addition to fish oil (Table S4 in Multimedia Appendix 1).

Figure 3. Histogram of child-reported quality of life scores.

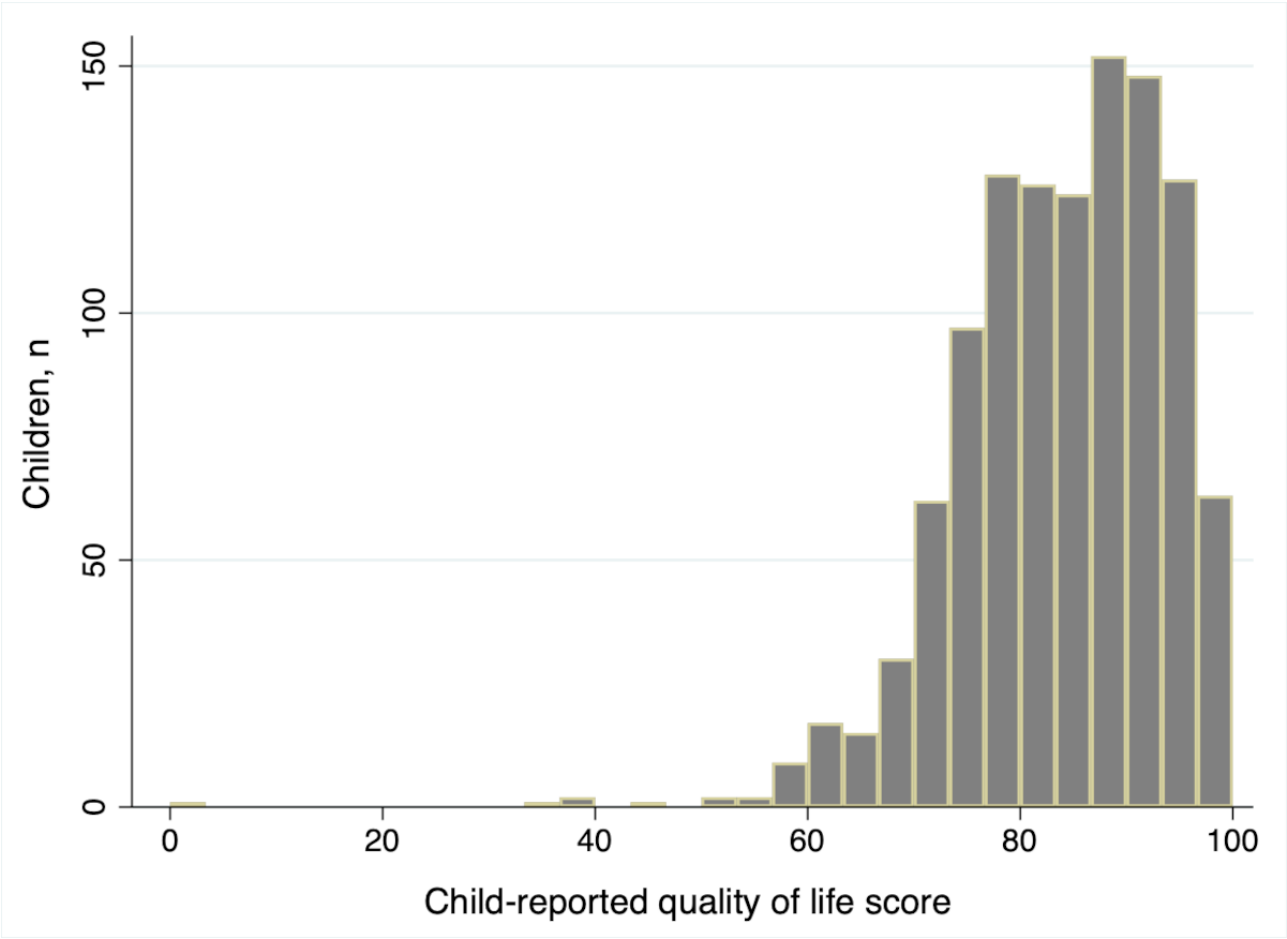


Table 6. Adjusted effects of with modifiable risk factors in mixed regression model for predicting quality of life.

Variable	β (95% CI)	n	*P* value
Frequency of physical activity (7-level quasi-continuous)	1.09 (0.53 to 1.66)	676	<.001
Hours of television per week (self-reported; 7-level quasi-continuous)	−0.95 (−1.55 to −0.36)	676	.002
Hard exercise (minutes)	0.02 (0.002 to 0.03)	676	.008
Percentage of time in moderate exercise	0.29 (0.002 to 0.59)	676	.048

Our parsimonious 3-variable mixed model explained 12% of variance in the training set and 15% of the variance in the validation set. Machine learning techniques retained more observations than the first regression model due to our selection of the fish oil variable, which had fewer observations (Table 7). Our second 4-predictor model explained 8% of the variance in the training set and 6% to 7% of the variance in the validation set. This was outperformed by support vector machine; however, our second regression model retained more observations and had been limited by us to modifiable risk factors.

XSL•FO
RenderX

Table 7. Performance indicators by modeling approach for quality of life.

Model	Training (n=1107)			Validation (n=453)		
	RMSE[a]	R^2 value	n	RMSE[a]	R^2 value	n
Regression model 1	0.89	0.11	293	0.85	0.13	111
Mixed model 1	0.89	0.12	293	0.85	0.15	111
Regression model 2	0.91	0.08	676	0.95	0.06	275
Mixed model 2	0.91	0.08	676	0.96	0.07	275
Random forest	0.66	0.74	481	0.89	0.03	190
Support vector machine	0.85	0.14	524	0.97	0.08	208
k-Nearest neighbors	0.78	0.33	295	0.97	0.08	117
Neural network	0.80	0.28	319	0.99	0.07	123

[a]RMSE: residual mean square error.

Table 8 shows the results from imputed sensitivity analyses. Regression models included 8 variables (Multimedia Appendix 1). The variables selected by the machine learning techniques are shown in Table S4 in Multimedia Appendix 1. The missing at random assumption was mostly acceptable, with 5 out of the 17 variables selected for modeling (hard exercise, percentage of time in moderate and light exercise, parent quality of life score, and master's education for father) having an effect on quality of life.

Table 8. Performance indicators by modeling approach for quality of life in sensitivity analysis (single-mean imputation).

Model	Training (n=1107)			Validation (n=453)		
	RMSE[a]	R^2 value	n	RMSE[a]	R^2 value	n
Regression model	0.95	.09	1107	0.93	.13	453
Mixed model	0.95	.09	1107	0.93	.14	453
Random forest	0.80	.59	1107	0.96	.05	453
Support vector machine	0.92	.17	1107	0.96	.07	453
k-Nearest neighbors	0.94	.12	1107	0.96	.06	453
Neural network	0.96	.09	1107	0.97	.05	453

[a]RMSE: residual mean square error.

Discussion

Principal Results and Comparisons to Existing Research

In modeling continuous health outcomes in a data set containing some missing data, linear regression was less prone to overfitting, retained more observations, and outperformed common machine learning techniques. In validation, regression explained approximately one-quarter of the variance in academic performance and up to 15% of the variance in quality of life, using exercise, lifestyle, and parental education quality of life data. Imputation improved machine learning performance, but improvements were not sufficient to outperform regression. Machine learning techniques outperformed regression for modeling nonlinear and heteroscedastic simulations and may be of use when there are no missing data or imputation is plausible, and where complex nonlinearity or heteroscedasticity exists. However, regression with splines performed almost as well for nonlinear modeling.

Multiple comparisons exist between machine learning techniques and logistic regression, multiclass, and survival analysis models, which taken together suggest similar results and an increased risk of overfitting with machine learning techniques [9,38-44]. However, few comparisons exist between machine learning techniques and linear regression for continuous health outcome measures. Hoffman et al [10] compared linear regression and support vector machine to predict Oswestry Disability Index score after surgery and found an adjusted R^2 of 0.42 for linear regression and 0.93 from support vector machine in a sample of 20 individuals. We observed that R^2 for support vector machine in our academic performance training set was approximately twice those for linear regression. However, the same relationship is not borne out in validation, suggesting the high R^2 value in the primary data is an artefact of overfitting. Laitinen and Räsänen [45] compared a regression equation with neural network in a sample of 125 patients with congenital heart disease and found that neural network performed best. However, the neural network used study data alone, and thus, was likely subject to overfitting, while the regression equation was externally validated. Hayward et al [11], in 91 patients with pancreatic cancer, compared linear regression to several machine learning techniques, including decision trees, k-nearest neighbors, and neural network across

XSL•FO
RenderX

a range of outcomes. They reported machine learning techniques and regression were comparable in 45 (35%) comparisons, machine learning techniques were superior in 33 (25%) comparisons, and machine learning techniques were inferior in 52 (40%) comparisons [11]. Our study uses more data than were used in prior work and more clearly demonstrates the superiority of linear regression for modeling continuous outcomes.

We found very strong evidence that reported physical activity, time recorded in vigorous exercise, and percentage of time spent in moderate exercise are positively associated with quality of life as continuous health outcomes in typical circumstances when adjusted for each of the other modeled variables. Associations between socioeconomic status, increased physical activity, and child quality of life are well established [13-15,46-48]. It has been suggested that the association may be explained via mechanisms involving affective response, increased self-efficacy, and improved mood-regulating neurotransmitter and endorphin release [14,49,50]. We found strong evidence that television and computer use is inversely proportional to quality of life. Increases of 1 use level (eg, going from 0 to 2 hours use per day), 100 minutes of vigorous exercise, or a 10% increase in exercise, are associated with small or small-to-medium (Multimedia Appendix 1) effects on quality of life. A systematic review [51] of physical activity and sedentary behavior on child quality of life found consistent evidence that watching television, using computers, or playing video games for more than 2 hours per day was significantly associated with lower child or adolescent quality of life. We found very strong evidence that parental assessment of child quality of life is associated with child quality of life assessment; this has been noted previously [25]. We found some evidence of association between academic performance and quality of life after adjustment; a 20-unit increase academic performance was associated with a small quality of life increase, and we are aware of no comparative work.

We found very strong evidence that reported physical activity, increased hand strength, mother having master's education or above, and decreased Stroop time, are associated with increases in academic performance. We found some evidence that a mother having university education and increases in television and computer use, are associated with increased academic performance. Reporting exercise that causes a sweat for at least 2 hours per week, 10 kg greater hand strength, a mother having university or master's education, increases of 1 television and computer use level, or a decreased Stroop time of 1 second were each associated with small or small-to-medium increases in academic performance. Socioeconomic status variables have been shown, in a meta-analysis [52] of 101,157 students, to be positively correlated with academic performance (with medium effect sizes), which is consistent with our findings. The role of socioeconomic status (ie, including parental education) may be explained by modified risk factors and health behaviors or self-concept [47,53]. Several mechanisms underlying a link between physical activity and academic performance have been suggested, which are thought to involve maintenance and facilitation of the plasticity of brain structures through altered neurogenesis and angiogenesis, enhanced central nervous system metabolism, and increased availability of growth factors [54-56]. An association between increasing physical activity and academic performance was demonstrated in a 2014 systematic review [57] of 215 studies. However, a 2019 systematic review [54] of 58 interventional studies of physical activity on cognitive performance, found only 10 out of 21 analyses (48%) in 5 high-quality studies demonstrated significant effects and found that the evidence was inconclusive. Furthermore, Singh et al [54] found only 15 of 25 analyses (60%) demonstrated academic performance benefits; stratification led to observation of strong evidence of a beneficial effect on math, but inconclusive evidence for language performance. Our own findings of an association between physical activity and general academic performance, come from using a composite outcome of reading, math, and English tests, and thus, future separate analyses may be of additional value.

Diet may affect both quality of life and academic performance via mechanisms related to the consumption of adequate micronutrients [17,58]. An association between healthy diet and the emotional functioning subscale of the Pediatric Quality of Life Inventory was demonstrated in a prospective study [18] of 3040 Australian adolescents (age 11 to 18 years). Our findings suggest small crude effects of diet across quality of life domains more generally. Decreased attendance, attention, and academic performance have been reported in undernourished children when compared to those reported in well-nourished children; fruit and vegetables, fat, and iron intake have been highlighted as having moderate effects in a study [58] of 5200 Canadian school children. A study [20] of 4245 Australian school-aged children (age 8-15 years) showed increased consumption of evening meal vegetables, breakfast consumption, and fruit are associated with higher spelling or writing scores, and increased sugar beverages are associated with lower scores. In our study, crude effects of increased sugared cordial consumption, sugar-free cordial, and pizza were associated with decreased academic performance generally but explained too little variance for us to select for inclusion in an adjusted model.

Implications

The rising popularity of machine learning techniques is understandable given the general abundance of data and a need for fewer assumptions. Machine learning techniques may be useful simply by virtue of the amount of data available. However, in public health research and health services research, data are less abundant and often missing. When modeling continuous outcomes in such circumstances, machine learning techniques are likely to perform worse unless marked nonlinear or heteroscedastic relationships exist. We have shown that the tendency to overfit that is often demonstrated in binary and multiclass machine learning techniques is also a challenge when modeling continuous outcomes. Furthermore, an innate inability for parameter estimation hampers interpretation and may make machine learning techniques generally less useful. At the time of writing, machine learning techniques have made relatively little impact in public health research on COVID-19 (with either continuous or categorical outcomes) where there is a pressing and immediate need for good modeling. We find this unsurprising—in most cases, public health data have normal distributions, and marked nonlinearity is rare. In these cases,

XSL•FO
RenderX

traditional regression methods use the most efficient estimators and will lead to better models.

Interventions aiming to improve activity levels in children may have a positive effect on both child quality of life and academic performance. The small association between academic performance and quality of life could follow satisfaction of achievement, although reversed causal direction, or residual confounding is plausible. In addition to increasing physical activity, new interventions to improve quality of life might target improvements in academic performance. Television and computer use is associated with decreases in quality of life but improvements in academic performance and these factors should be examined separately to clarify other promising intervention targets.

Strengths, Limitations, and Recommendations for Future Research

We provide like-for-like comparisons between machine learning techniques and regression for modeling continuous health outcomes, with larger sample size than those used in previous research, and separate validation. Nevertheless, our work has limitations. We used an average of reading, math, and English tests as a proxy for academic performance. Not including subjects such as science may impair construct coverage of academic performance. Using single-mean imputation and last observation carried forward (in missing Ungkost variables) allowed us to avoid using multiple imputation (which is based on regression approaches) for data used in machine learning models (ie, to avoid mixing methods). However, multiple imputation provides better coverage than single-mean imputation, and last observation carried forward is known to be problematic [59]. It has been highlighted that the assumption of no change over (limited) time may hold in some contexts and can be better than ignoring missingness altogether [60]. In our case, we believed the assumption of no or limited change would be better than ignoring missingness completely or mixing methods when comparing regression approaches with machine learning techniques. There is a potential limitation regarding the validity and generalizability of results to 11- and 12-year-old children in the case of greater than assumed unobserved changes in missing Ungkost variables. With respect to single-mean imputation, our results showed that the missing at random assumption was not valid for some modeled variables. We believe that the applied techniques have been kept robust to imputation issues because results were in alignment with those from complete-case analyses; however, results derived from our imputed sensitivity analyses should be interpreted cautiously. Generalization of results to other countries should also be done with caution, since there may be baseline differences in activity and culture among Norwegian children. Finally, we focused on machine learning techniques that we judged to be the most common and which we thought researchers would find useful; we acknowledge that this is not a comprehensive comparison of regression with all possible machine learning techniques.

Future focus on comparisons to other machine learning techniques, separate analysis of academic performance components, and iteratively varying the size of the training set to explore how training set size affects overfitting will provide further useful knowledge. The Ungkost item on television and computer use combines 2 activities. We found large positive associations between the item and academic performance and a small negative association with quality of life. We suspect the positive associations may be grounded in computer use for education, and the negative associations may be grounded in uses for leisure. Separation of these exposures will provide clarity. Some machine learning techniques retained diet variables that we did not select for adjusted models. One strength of machine learning techniques may be an ability to detect mild and easily missed nonlinear relationships, which is worth further exploration.

Conclusions

For modeling continuous health outcomes when some data are missing, linear regression is less prone to overfitting and outperforms common machine learning techniques. Imputation improves the performance of machine learning techniques, but improvements are not sufficient to outperform regression. Machine learning techniques outperform regression in modeling nonlinear and heteroscedastic relationships and may be of use in cases where imputation is sensible or there are no or few missing data. Otherwise regression is preferred. Regression with splines performs almost as well in nonlinear modeling. Lifestyle variables, including physical activity, television and computer use, muscular strength, and parental education were predictive of academic performance or quality of life explaining up to 24% and 15% of the variance in these outcomes, respectively. Targeting these areas in future interventions may help improve child quality of life and academic performance.

Acknowledgments

Thanks are due to Kristiania University College for providing seed funding for this work and to Gary Abel (University of Exeter), Sandra Eldridge (Queen Mary, University of London), and George Bouliotis (University of Warwick) for helpful discussions related to this work.

Authors' Contributions

RF conceived the study, applied for internal seed funding, conducted some analyses, and wrote the first draft of the manuscript. SH conducted most of the machine learning analyses, and HK conducted remaining machine learning analyses. JF set up and maintained study software and server. LF provided input on educational components. PMF provided data and input on the HOPP

XSL•FO
RenderX

study and obtained ethics approval for the HOPP study activities. All authors contributed to interpretation of the findings and approved the final manuscript.

Conflicts of Interest

RF is a director and shareholder and JF is a shareholder of Clinvivo Ltd, a University of Warwick spin-out company. Neither Clinvivo services nor Clinvivo software products were used in this study.

Multimedia Appendix 1
Supplementary tables and technical notes.
[PDF File (Adobe PDF File), 472 KB - jmir_v23i7e22021_app1.pdf]

References

1. Froud R, Rajendran D, Patel S, Bright P, Bjørkli T, Eldridge S, et al. The power of low back pain trials: a systematic review of power, sample size, and reporting of sample size calculations over time, in trials published between 1980 and 2012. Spine (Phila Pa 1976) 2017 Jun 01;42(11):E680-E686. [doi: 10.1097/BRS.0000000000001953] [Medline: 27792111]
2. Froud R, Patel S, Rajendran D, Bright P, Bjørkli T, Buchbinder R, et al. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. PLoS One 2016;11(10):e0164573 [FREE Full text] [doi: 10.1371/journal.pone.0164573] [Medline: 27776141]
3. Bhatia A, Yu-Wei C. Machine Learning With R Cookbook 2nd edition. Birmingham: Pact Publishing; 2017.
4. Michie D, Spiegelhalter DC. Machine Learning, Neural and Statistical Classification. Leeds: University of Leeds; 1994.
5. Lötsch J, Ultsch A. Machine learning in pain research. Pain 2018 Apr;159(4):623-630 [FREE Full text] [doi: 10.1097/j.pain.0000000000001118] [Medline: 29194126]
6. Diller G, Kempny A, Babu-Narayan SV, Henrichs M, Brida M, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. Eur Heart J 2019 Apr 01;40(13):1069-1077 [FREE Full text] [doi: 10.1093/eurheartj/ehy915] [Medline: 30689812]
7. Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised machine learning to identify high likelihood of dementia in population-based surveys: development and validation study. J Med Internet Res 2018 Dec 09;20(7):e10493 [FREE Full text] [doi: 10.2196/10493] [Medline: 29986849]
8. Wellner B, Grand J, Canzone E, Coarr M, Brady PW, Simmons J, et al. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. JMIR Med Inform 2017 Nov 22;5(4):e45 [FREE Full text] [doi: 10.2196/medinform.8680] [Medline: 29167089]
9. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer 2001 Apr 15;91(8 Suppl):1636-1642. [doi: 10.1002/1097-0142(20010415)91:8+<1636::aid-cncr1176>3.0.co;2-d] [Medline: 11309761]
10. Hoffman H, Lee SI, Garst JH, Lu DS, Li CH, Nagasawa DT, et al. Use of multivariate linear regression and support vector regression to predict functional outcome after surgery for cervical spondylotic myelopathy. J Clin Neurosci 2015 Sep;22(9):1444-1449 [FREE Full text] [doi: 10.1016/j.jocn.2015.04.002] [Medline: 26115898]
11. Hayward J, Alvarez SA, Ruiz C, Sullivan M, Tseng J, Whalen G. Machine learning of clinical performance in a pancreatic cancer database. Artif Intell Med 2010 Jul;49(3):187-195. [doi: 10.1016/j.artmed.2010.04.009] [Medline: 20483571]
12. Solans M, Pane S, Estrada M, Serra-Sutton V, Berra S, Herdman M, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. Value Health 2008;11(4):742-764 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00293.x] [Medline: 18179668]
13. Ringdal K, Ringdal GI, Olsen HK, Mamen A, Fredriksen PM. Quality of life in primary school children: the Health Oriented Pedagogical Project (HOPP). Scand J Public Health 2018 May;46(21_suppl):68-73. [doi: 10.1177/1403494818767821] [Medline: 29754578]
14. Moeijes J, van Busschbach JT, Bosscher RJ, Twisk JWR. Sports participation and health-related quality of life: a longitudinal observational study in children. Qual Life Res 2019 Sep;28(9):2453-2469 [FREE Full text] [doi: 10.1007/s11136-019-02219-4] [Medline: 31161332]
15. Moeijes J, van Busschbach JT, Wieringa TH, Kone J, Bosscher RJ, Twisk JWR. Sports participation and health-related quality of life in children: results of a cross-sectional study. Health Qual Life Outcomes 2019 Apr 15;17(1):64 [FREE Full text] [doi: 10.1186/s12955-019-1124-y] [Medline: 30987637]
16. Jozefiak T, Sønnichsen Kayed N. Self- and proxy reports of quality of life among adolescents living in residential youth care compared to adolescents in the general population and mental health services. Health Qual Life Outcomes 2015 Jul 22;13:104 [FREE Full text] [doi: 10.1186/s12955-015-0280-y] [Medline: 26197764]
17. O'Neil A, Quirk SE, Housden S, Brennan SL, Williams LJ, Pasco JA, et al. Relationship between diet and mental health in children and adolescents: a systematic review. Am J Public Health 2014 Oct;104(10):e31-e42. [doi: 10.2105/AJPH.2014.302110] [Medline: 25208008]

XSL•FO
RenderX

18. Jacka FN, Kremer PJ, Berk M, de Silva-Sanigorski AM, Moodie M, Leslie ER, et al. A prospective study of diet quality and mental health in adolescents. PLoS One 2011;6(9):e24805 [FREE Full text] [doi: 10.1371/journal.pone.0024805] [Medline: 21957462]
19. Hughes AR, Farewell K, Harris D, Reilly JJ. Quality of life in a clinical sample of obese children. Int J Obes (Lond) 2007 Jan;31(1):39-44. [doi: 10.1038/sj.ijo.0803410] [Medline: 16733522]
20. Burrows T, Goldman S, Olson RK, Byrne B, Coventry WL. Associations between selected dietary behaviours and academic achievement: a study of Australian school aged children. Appetite 2017 Sep 01;116:372-380. [doi: 10.1016/j.appet.2017.05.008] [Medline: 28487248]
21. Ren X, Schweizer K, Wang T, Xu F. The prediction of students' academic performance with fluid intelligence in giving special consideration to the contribution of learning. Adv Cogn Psychol 2015;11(3):97-105 [FREE Full text] [doi: 10.5709/acp-0175-z] [Medline: 26435760]
22. Sohr-Preston SL, Scaramella LV, Martin MJ, Neppl TK, Ontai L, Conger R. Parental socioeconomic status, communication, and children's vocabulary development: a third-generation test of the family investment model. Child Dev 2013;84(3):1046-1062 [FREE Full text] [doi: 10.1111/cdev.12023] [Medline: 23199236]
23. Fredriksen PM, Hjelle OP, Mamen A, Meza TJ, Westerberg AC. The Health Oriented Pedagogical Project (HOPP) - a controlled longitudinal school-based physical activity intervention program. BMC Public Health 2017 Apr 28;17(1):370 [FREE Full text] [doi: 10.1186/s12889-017-4282-z] [Medline: 28454531]
24. Jozefiak T, Mattejat F, Remschmidt H. Inventory of Life Quality in Children and Adolescents Manual, Norwegian version. Stockholm, Sweden: Hogrefe; 2012.
25. Jozefiak T, Larsson B, Wichstrøm L, Mattejat F, Ravens-Sieberer U. Quality of life as reported by school children and their parents: a cross-sectional survey. Health Qual Life Outcomes 2008 May 19;6:34. [doi: 10.1186/1477-7525-6-34] [Medline: 18489777]
26. Bjørnsson J. Metodegrunnlag for nasjonale prøver [Methodological basis for national tests]. Oslo: Utdanningsdirektoratet; 2018.
27. Deng WH, Fredriksen PM. Objectively assessed moderate-to-vigorous physical activity levels among primary school children in Norway: the Health Oriented Pedagogical Project (HOPP). Scand J Public Health 2018 May;46(21_suppl):38-47. [doi: 10.1177/1403494818771207] [Medline: 29754576]
28. Andersen LB, Andersen TE, Andersen E, Anderssen SA. An intermittent running test to estimate maximal oxygen uptake: the Andersen test. J Sports Med Phys Fitness 2008 Dec;48(4):434-437. [Medline: 18997644]
29. Stroop J. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18:643-662 [FREE Full text]
30. Scarpina F, Tagini S. The Stroop color and word test. Front Psychol 2017;8:557 [FREE Full text] [doi: 10.3389/fpsyg.2017.00557] [Medline: 28446889]
31. Øverby N, Andersen L. Ungkost-2000. Landsomfattende kostholdsundersøkelse blant elver i 4. og 8. klasse i Norge [Nationwide dietary survey among pupils in 4th and 8th grade in Norway]. Oslo: Sosial og helsedirektoratet; 2002.
32. Gould W. Linear splines and piecewise linear functions. Stata Technical Bulletin 1993;5:13-17 [FREE Full text]
33. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. PLoS One 2014;9(12):e114468 [FREE Full text] [doi: 10.1371/journal.pone.0114468] [Medline: 25474472]
34. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974 Dec;19(6):716-723. [doi: 10.1109/TAC.1974.1100705]
35. James G, Witten D, Hastie T, Tibshirabi R. An Introduction to Statistical Learning With Applications in R. New York: Springer; 2013.
36. Bonaccorso G. Machine Learning Algorithms second edition. Birmingham: Packt; 2020.
37. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
38. Tu J. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996 Nov;49(11):1225-1231. [doi: 10.1016/s0895-4356(96)00002-9] [Medline: 8892489]
39. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA Netw Open 2020 Jan 03;3(1):e1918962 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18962] [Medline: 31922560]
40. Lorenzoni G, Sabato SS, Lanera C, Bottigliengo D, Minto C, Ocagli H, et al. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. J Clin Med 2019 Aug 24;8(9):1298 [FREE Full text] [doi: 10.3390/jcm8091298] [Medline: 31450546]
41. Salcedo-Bernala A, Villamil-Giraldoa M, Moreno-Barbosaa A. Clinical data analysis: an opportunity to compare machine learning methods. Procedia Comput Sci 2016;100:731-738 [FREE Full text]
42. Faisal M, Scally A, Howes R, Beatson K, Richardson D, Mohammed MA. A comparison of logistic regression models with alternative machine learning methods to predict the risk of in-hospital mortality in emergency medical admissions via external validation. Health Informatics J 2020 Mar;26(1):34-44. [doi: 10.1177/1460458218813600] [Medline: 30488755]

XSL•FO
RenderX

43. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen AC, Joseph KS, et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. BMC Pregnancy Childbirth 2018 Aug 15;18(1):333 [FREE Full text] [doi: 10.1186/s12884-018-1971-2] [Medline: 30111303]
44. Yahya N, Ebert MA, Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Statistical-learning strategies generate only modestly performing predictive models for urinary symptoms following external beam radiotherapy of the prostate: a comparison of conventional and machine-learning methods. Med Phys 2016 May;43(5):2040. [doi: 10.1118/1.4944738] [Medline: 27147316]
45. Laitinen PO, Räsänen J. Measured versus predicted oxygen consumption in children with congenital heart disease. Heart 1998 Dec;80(6):601-605 [FREE Full text] [doi: 10.1136/hrt.80.6.601] [Medline: 10065031]
46. Jirojanakul P, Skevington SM, Hudson J. Predicting young children's quality of life. Soc Sci Med 2003 Oct;57(7):1277-1288. [doi: 10.1016/s0277-9536(02)00507-5] [Medline: 12899910]
47. von Rueden U, Gosch A, Rajmil L, Bisegger C, Ravens-Sieberer U. Socioeconomic determinants of health related quality of life in childhood and adolescence: results from a European study. J Epidemiol Community Health 2006 Feb;60(2):130-135 [FREE Full text] [doi: 10.1136/jech.2005.039792] [Medline: 16415261]
48. Marques A, Mota J, Gaspar T, de Matos MG. Associations between self-reported fitness and self-rated health, life-satisfaction and health-related quality of life among adolescents. J Exerc Sci Fit 2017 Jun;15(1):8-11 [FREE Full text] [doi: 10.1016/j.jesf.2017.03.001] [Medline: 29541124]
49. Rhodes RE, Kates A. Can the affective response to exercise predict future motives and physical activity behavior? a systematic review of published evidence. Ann Behav Med 2015 Oct;49(5):715-731. [doi: 10.1007/s12160-015-9704-5] [Medline: 25921307]
50. García-Hermoso A, Hormazábal-Aguayo I, Fernández-Vergara O, Olivares PR, Oriol-Granado X. Physical activity, screen time and subjective well-being among children. Int J Clin Health Psychol 2020;20(2):126-134 [FREE Full text] [doi: 10.1016/j.ijchp.2020.03.001] [Medline: 32550852]
51. Wu XY, Han LH, Zhang JH, Luo S, Hu JW, Sun K. The influence of physical activity, sedentary behavior on health-related quality of life among the general population of children and adolescents: a systematic review. PLoS One 2017;12(11):e0187668 [FREE Full text] [doi: 10.1371/journal.pone.0187668] [Medline: 29121640]
52. Sirin S. Socioeconomic status and academic achievement: a meta-analytic review of research. Rev Educ Res 2005;75(3):417-453 [FREE Full text]
53. Li S, Xu Q, Xia R. Relationship between SES and academic achievement of junior high school students in China: the mediating effect of self-concept. Front Psychol 2019;10:2513. [doi: 10.3389/fpsyg.2019.02513] [Medline: 31956315]
54. Singh AS, Saliasi E, van den Berg V, Uijtdewilligen L, de Groot RHM, Jolles J, et al. Effects of physical activity interventions on cognitive and academic performance in children and adolescents: a novel combination of a systematic review and recommendations from an expert panel. Br J Sports Med 2019 May;53(10):640-647. [doi: 10.1136/bjsports-2017-098136] [Medline: 30061304]
55. Cotman CW, Berchtold NC, Christie L. Exercise builds brain health: key roles of growth factor cascades and inflammation. Trends Neurosci 2007 Sep;30(9):464-472. [doi: 10.1016/j.tins.2007.06.011] [Medline: 17765329]
56. van Praag H. Neurogenesis and exercise: past and future directions. Neuromolecular Med 2008;10(2):128-140. [doi: 10.1007/s12017-008-8028-z] [Medline: 18286389]
57. Castelli DM, Centeio EE, Hwang J, Barcelona JM, Glowacki EM, Calvert HG, et al. VII. The history of physical activity and academic performance research: informing the future. Monogr Soc Res Child Dev 2014 Dec;79(4):119-148. [doi: 10.1111/mono.12133] [Medline: 25387418]
58. Florence MD, Asbridge M, Veugelers PJ. Diet quality and academic performance. J Sch Health 2008 Apr;78(4):209-15; quiz 239. [doi: 10.1111/j.1746-1561.2008.00288.x] [Medline: 18336680]
59. Vickers AJ, Altman DG. Statistics notes: missing outcomes in randomised trials. BMJ 2013 Jun 06;346:f3438. [doi: 10.1136/bmj.f3438] [Medline: 23744649]
60. Shoop SJW. Should we ban the use of 'last observation carried forward' analysis in epidemiological studies? SM J Public Health Epidemiol 2015;1(1):1004 [FREE Full text]

Abbreviations

COVID-19: coronavirus disease 2019
HOPP: Health Oriented Pedagogical Project
RMSE: residual mean square error

XSL•FO
RenderX

Edited by R Kukafka, G Eysenbach; submitted 03.07.20; peer-reviewed by T Wieringa, A Benis; comments to author 27.09.20; revised version received 26.10.20; accepted 17.05.21; published 16.07.21.

Please cite as:
Froud R, Hansen SH, Ruud HK, Foss J, Ferguson L, Fredriksen PM
Relative Performance of Machine Learning and Linear Regression in Predicting Quality of Life and Academic Performance of School Children in Norway: Data Analysis of a Quasi-Experimental Study
J Med Internet Res 2021;23(7):e22021
URL: https://www.jmir.org/2021/7/e22021
doi:10.2196/22021
PMID:34009128

©Robert Froud, Solveig Hakestad Hansen, Hans Kristian Ruud, Jonathan Foss, Leila Ferguson, Per Morten Fredriksen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Studying Microtemporal, Within-Person Processes of Diet, Physical Activity, and Related Factors Using the APPetite-Mobile-App: Feasibility, Usability, and Validation Study

Alea Ruf[1], BSc, MSc; Elena Doris Koch[2], BA, MA; Ulrich Ebner-Priemer[2,3], BA, MA, PhD, Prof Dr; Monika Knopf[4], BA, MA, PhD, Prof Dr; Andreas Reif[1], Dr med, Prof Dr; Silke Matura[1], BA, MA, Dr rer med

[1]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Goethe University, Frankfurt, Germany

[2]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[3]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[4]Department of Developmental Psychology, Goethe University, Frankfurt, Germany

Corresponding Author:
Alea Ruf, BSc, MSc
Department of Psychiatry, Psychosomatic Medicine and Psychotherapy
University Hospital
Goethe University
Heinrich-Hoffmann-Straße 10
Frankfurt, 60528
Germany
Phone: 49 69 6301 83348
Email: alea.ruf@kgu.de

Abstract

Background: Diet and physical activity (PA) have a major impact on physical and mental health. However, there is a lack of effective strategies for sustaining these health-protective behaviors. A shift to a microtemporal, within-person approach is needed to capture dynamic processes underlying eating behavior and PA, as they change rapidly across minutes or hours and differ among individuals. However, a tool that captures these microtemporal, within-person processes in daily life is currently not present.

Objective: The APPetite-mobile-app is developed for the ecological momentary assessment of microtemporal, within-person processes of complex dietary intake, objectively recorded PA, and related factors. This study aims to evaluate the feasibility and usability of the APPetite-mobile-app and the validity of the incorporated APPetite-food record.

Methods: The APPetite-mobile-app captures dietary intake event-contingently through a food record, captures PA continuously through accelerometers, and captures related factors (eg, stress) signal-contingently through 8 prompts per day. Empirical data on feasibility (n=157), usability (n=84), and validity (n=44) were collected within the Eat2beNICE-APPetite-study. Feasibility and usability were examined in healthy participants and psychiatric patients. The relative validity of the APPetite-food record was assessed with a subgroup of healthy participants by using a counterbalanced crossover design. The reference method was a 24-hour recall. In addition, the energy intake was compared with the total energy expenditure estimated from accelerometry.

Results: Good feasibility, with compliance rates above 80% for prompts and the accelerometer, as well as reasonable average response and recording durations (prompt: 2.04 min; food record per day: 17.66 min) and latencies (prompts: 3.16 min; food record: 58.35 min) were found. Usability was rated as moderate, with a score of 61.9 of 100 on the System Usability Scale. The evaluation of validity identified large differences in energy and macronutrient intake between the two methods at the group and individual levels. The APPetite-food record captured higher dietary intakes, indicating a lower level of underreporting, compared with the 24-hour recall. Energy intake was assessed fairly accurately by the APPetite-food record at the group level on 2 of 3 days when compared with total energy expenditure. The comparison with mean total energy expenditure (2417.8 kcal, SD 410) showed that the 24-hour recall (1909.2 kcal, SD 478.8) underestimated habitual energy intake to a larger degree than the APPetite-food record (2146.4 kcal, SD 574.5).

Conclusions: The APPetite-mobile-app is a promising tool for capturing microtemporal, within-person processes of diet, PA, and related factors in real time or near real time and is, to the best of our knowledge, the first of its kind. First evidence supports the good feasibility and moderate usability of the APPetite-mobile-app and the validity of the APPetite-food record. Future

XSL•FO
RenderX

findings in this context will build the foundation for the development of personalized lifestyle modification interventions, such as just-in-time adaptive interventions.

(J Med Internet Res 2021;23(7):e25850) doi:10.2196/25850

KEYWORDS

diet; physical activity; microtemporal processes; within-person factors; ecological momentary assessment; smartphone-app; mobile phone; mHealth; dietary assessment; feasibility; usability; validity

Introduction

Background

Diet is a key contributor to both physical and mental health. Elevated BMI is a major risk factor for noncommunicable diseases, such as cardiovascular diseases [1]. Since 1975, the prevalence of obesity has nearly tripled globally [1]. Accordingly, in 2016, 13% of adults were obese and 39% were overweight [1]. Approximately 11 million deaths were associated with dietary risk factors (eg, low intake of whole grains) across 195 countries in 2017 [2]. Although the link between diet and mental health is not equally well understood, first evidence supports the presence of a direct association among diet, mental health, and mental functioning [3]. Obesity not only increases the probability of somatic diseases but also of mental illness, particularly depression [4-6]. These numbers and findings highlight the growing need to understand the "causes of the causes."

Although factors and processes underlying eating behavior have been studied for many years [7,8], interventions remain ineffective in sustaining health-protective behaviors for the long term [9]. One reason for this could be the main focus on between-person characteristics (eg, age) and macrotemporal processes (across weeks, months, or years) [10]. Diet is a highly complex health behavior that is performed multiple times per day and is influenced by a variety of fluctuating factors and their interactions [11]. A real-life microtimescale approach is needed to capture the dynamics of diet and associated factors ecologically and momentarily and, ultimately, to understand the processes underlying eating behavior in everyday life [10]. In contrast to some between-person characteristics (eg, age), within-person factors are modifiable and therefore a promising target for interventions. For this reason, the identification of within-person factors that influence eating behavior in daily life is needed for the development of novel, more effective, and personalized interventional approaches.

It is not only diet that has a large impact on both physical and mental health. Physical activity (PA) represents another impactful, repeated-occurrence health behavior [12,13]. To untangle the complex association between diet and health [14], it is important to consider possible interactions. For instance, diet does not independently regulate body weight. Body weight is regulated through the interplay of energy intake (ie, diet) and energy expenditure (eg, PA) [15]. Therefore, the assessment of microtemporal, within-person processes of diet and PA should be combined, and possible interactions should be taken into consideration.

Ecological Momentary Assessment of Diet, PA, and Related Factors

The repeated or continuous assessment of experiences, behaviors, or physiological processes in real life through smartphones or wearable devices is a highly promising approach for studying microtemporal, within-person processes [16]. This approach is referred to as ecological momentary assessment (EMA), ambulatory assessment, experience sampling, and real-time data capture [17]. Although different terms have been used, they have in common the assessment of various parameters, multiple times per day in daily life [17].

Even though EMA studies do not allow causal conclusions, they offer insight into three important aspects of microtemporal, within-person processes: (1) temporal specificity (eg, Does diet influence mood to a greater extent than mood influences diet?), (2) situational specificity (eg, Is unhealthy eating more likely when being alone or with others?), and (3) person specificity (eg, Is stress more predictive for engaging in eating for some individuals compared with others?) [10].

Diet is a highly complex phenomenon that makes its assessment difficult. However, to avoid typical reporting biases that are present in traditional dietary assessment methods (eg, food frequency questionnaires), the number of studies using EMA to capture self-reported dietary intake or aspects of it in real time or near real time instead of retrospectively has rapidly grown in the last decade [18-20]. There are two categories of EMA approaches present so far: on the one hand, there are mobile-based dietary assessment tools that focus on the assessment of complex dietary intake and the generation of nutritional values. Complex dietary intake refers to assessing all consumed foods and drinks and consumed amounts, which are then used to generate nutritional values. Even though a small number of tools that assess complex dietary intake also allow assessing contextual correlates during eating occasions [21], no tool allows capturing a wider repertoire of factors preceding or succeeding eating occasions [22]. On the other hand, there are a number of studies that use EMA to study a variety of factors related to diet (eg, affect [23]). However, to the best of our knowledge, none of these studies assessed diet in its full complexity. Most of them focus on specific aspects of diet only, for example, snacks or sweetened beverages [24-29], a limited number of food and drink categories [23,30-33], portion sizes [34], or the type of eating events (main meals vs snacks) and the type of drinking occasions (alcoholic vs nonalcoholic) [35]. Hence, complex dietary intake was not assessed, and the generation of nutritional values was not possible. Although some of these studies reported a more comprehensive approach which captured all consumed foods and, in some studies, drinks through a free input field [23,33], the foods and drinks were

XSL•FO
RenderX

only assigned to a limited number of food and drink categories and were not used to generate nutritional values. There is a need to study the processes underlying complex dietary intake instead of processes underlying only aspects of diet.

Despite the importance of taking possible interactions into account, most EMA studies focus on either the assessment of diet or PA. One study identified the need for an EMA tool to capture *complex lifestyle behavior*, that is, dietary intake and PA simultaneously [36]. However, the tool developed for this purpose failed to assess diet and PA in their complex nature. It only assessed specific food categories and used self-reports for the assessment of PA, which is unsatisfactory, given that 2 systematic reviews showed that indirect measures of PA (ie, self-reports) differ substantially from direct, objective measures (eg, accelerometers) [37,38].

In conclusion, there is a strong need for an EMA tool that allows capturing complex dietary intake, objectively measured PA, and a broad range of associated factors simultaneously in daily life to study microtemporal, within-person processes underlying these health-protective behaviors.

Objectives

As no EMA tool allows the study of microtemporal, within-person processes of complex dietary intake, objectively measured PA, and related factors, we developed an EMA tool for the simultaneous assessment of these complex health behaviors and related factors in daily life: the APPetite-mobile-app (this term also covers the assessment of PA, although it is not performed by the APPetite-mobile-app itself but by an accelerometer).

The suitability of novel EMA tools for use in daily life should be evaluated. Therefore, feasibility, usability, and validity were examined empirically in this study. The following questions will be addressed: Is the APPetite-mobile-app a feasible and usable tool for the combined assessment of complex dietary intake, PA, and associated factors in daily life and a valid tool for the assessment of complex dietary intake in real time or near real time?

Methods

The APPetite-Mobile-App

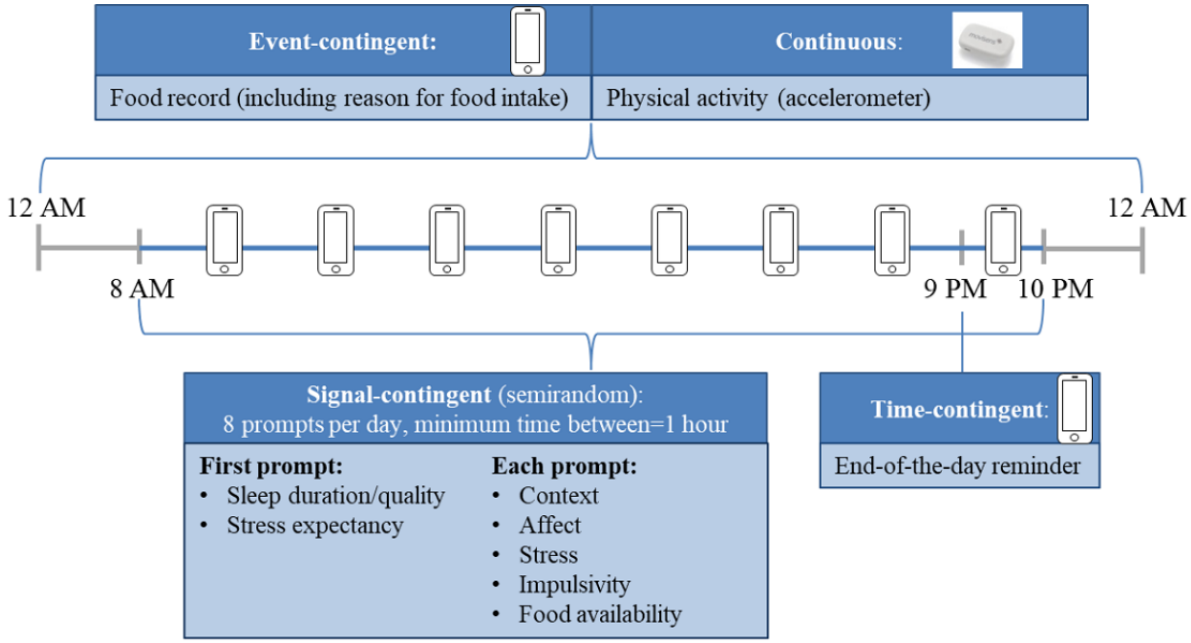
Software and Hardware

The APPetite-mobile-app was developed and run through movisensXS (version 1.4.7, movisens GmbH), a web-based platform for the development of EMA tools. It supports a broad range of sampling schemes, item formats, and multimedia records, allowing flexible and tailored study configurations. The APPetite-mobile-app is run through the movisensXS app (available for Android devices). If the mobile device has access to mobile data during the EMA assessment, participants' entries will be uploaded instantly to the platform. In this way, compliance can be monitored throughout the EMA assessment, and a chat function allows direct messaging with participants. All participants received a study smartphone (Motorola Moto G 3rd generation), with access to mobile data. The movisensXS app was previously tested on this particular mobile device, ensuring its smooth functioning and increasing the standardization of the mobile-based assessment.

Sampling Strategy

The APPetite-mobile-app uses event-, signal-, and time-contingent as well as continuous sampling (Figure 1).

Figure 1. Sampling strategy of the APPetite-mobile-app.



Food intake was recorded event-contingently through a food record. Participants were asked to enter foods and drinks as soon as possible after consuming them. Accordingly, participants are able to initiate the APPetite-food record at any time and capture their food intake in real time. This was chosen to minimize memory effects and record the exact time of food

XSL•FO
RenderX

intake. In addition, this allows capturing food intake even during the night, when signal-contingent prompts are inappropriate. At 9 PM, a time-contingent prompt asks if all consumed foods and drinks of the day have been recorded, ensuring that no foods and drinks consumed on this day are missed.

The prompts are initiated signal-contingent at eight semirandom times per day between 8 AM and 10 PM. The minimum time between 2 prompts is 1 hour. Therefore, participants cannot predict the exact time of the next prompt, and the assessed situation is a better reflection of the participant's real life. Participants were instructed to respond immediately to the prompt. However, if participants are unable to reply instantly, it is possible to postpone the prompt for 5, 10, 15, 20, or 25 minutes to avoid missing data and reduce the participants' burden. If no reaction is registered, the prompt is deactivated and cannot be reactivated.

Continuous sampling through an accelerometer is used for the assessment of PA.

EMA Measures

APPetite-Food Record

The APPetite-food record comprises a 6-step process: (1) selection of meal type, (2) entry of time of intake, (3) selection of consumed foods and drinks, (4) specification of consumed amounts, (5) presentation of reminder for commonly forgotten foods, and (6) indication of predominant reason for eating or drinking (Figure 2 presents screenshots of the 6-step process). To generate nutritional values, the obtained dietary data were transferred by trained staff to myfood24-Germany, a 24-hour dietary recall [39]. A detailed description of the APPetite-food record and nutritional value generation is provided in Multimedia Appendix 1 [39-44]. All reasons for eating and drinking are presented in Multimedia Appendix 2 [40-44].

Figure 2. Screenshots of the 6-step process of the APPetite-food record.



Prompts

Each prompt assesses the context, affect, stress, impulsivity, and food availability either since the last prompt or immediately before the prompt. In addition, the first prompt of a day captures sleep quality and quantity as well as stress expectancy. All prompt measures and items are described in Multimedia Appendix 3 [45-49].

Physical Activity

Move 3 sensors from movisens were used to objectively record PA. The accelerometer was worn on the nondominant wrist.

XSL•FO
RenderX

Participants were asked to wear it at any time (also when sleeping) and only take it off when showering or performing water activities. The Move 3 sensor captures raw data on 3D acceleration, barometric air pressure, and temperature. Secondary parameters such as activity class, body position, steps, metabolic equivalents, and PA metrics can be extracted using the DataAnalyzer (movisens GmbH).

Evaluation of Feasibility

Measures

The feasibility of the APPetite-mobile-app was separately assessed for the EMA prompts, the APPetite-food record, and the accelerometer. The feasibility of the prompts is determined by prompt delivery, total number of answered prompts across all subjects, number of answered prompts per participant, compliance (percentage of complete prompts within received prompts), response latency (time from first prompt signal to answering), and the time needed to complete a single prompt. The food record's feasibility was evaluated based on the number of recorded eating and drinking events per day, reporting latency (time between the meal and meal recording), and the time needed to record food intake per day. The amount of time wearing the accelerometer and compliance (percentage wearing the accelerometer within the 7-day assessment period) are measures of feasibility of the accelerometer.

Sample

The data were collected within the ongoing APPetite study. The APPetite study is part of the European Union Horizon2020 project Eat2beNICE and recruits participants from three existing studies: LORA (Longitudinal Resilience Assessment) study [50], PROUD (Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder) study [51], and the BipoLife-A1 study (improving early recognition and intervention in people at risk of developing bipolar disorder [52,53]). The LORA study included individuals who were not affected by psychiatric conditions. The PROUD sample consisted of patients affected by attention-deficit/hyperactivity disorder. The BipoLife-A1 study follows up on patients with an increased risk for the development of bipolar disorder, including patients affected by attention-deficit/hyperactivity disorder or depression.

From November 2018 to March 2020, 161 participants were included in the APPetite study (140 LORA, 7 PROUD, and 14 BipoLife-A1). After the first in-person session, 3 LORA participants dropped out. Of these, 2 realized that they were unable to respond to prompts. The third person was mistakenly given a smartphone that was not coupled with the EMA protocol. Another person dropped out after the additional in-person session of the validation study for private reasons. Hence, EMA data of 157 participants are available for the evaluation of feasibility (see demographics in Table 1).

Table 1. Demographics of the total sample and the 3 cohorts (only individuals who completed the ecological momentary assessment were included; N=157).

Variables	Total (N=157)	LORA[a] (n=136)	PROUD[b] (n=7)	BipoLife-A1 (n=14)
Gender, n (%)				
Female	100 (63.7)	94 (69.1)	1 (14.3)	5 (35.7)
Male	57 (36.3)	42 (30.9)	6 (85.7)	9 (64.3)
Age (years), mean (SD)	28.04 (7.22)	28.08 (7.55)	26.43 (2.51)	28.5 (5.39)
BMI, mean (SD)	24.71 (4.81)	24.26 (3.87)	24.98 (6.11)	28.9 (9.13)

[a]LORA: Longitudinal Resilience Assessment.

[b]PROUD: Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder.

Procedure

The APPetite study consists of 2 in-person sessions, the EMA assessment, and a follow-up session from home. In the first in-person session, participants received detailed training on how to use the APPetite-mobile-app and the accelerometer. Participants received a smartphone with the APPetite-mobile-app and an accelerometer, including a wristband. Participants used the APPetite-mobile-app for 3 consecutive days (2 weekdays and 1 weekend day, not including the day of the first in-person session) and wore the accelerometer for 7 consecutive days (overlapping the 3 days of the APPetite-mobile-app assessment, not including the day of the first in-person session). During the 3 days of the app-based assessment, prompt compliance was tracked. If compliance fell below the threshold of 80%, a motivational message was sent to the participant. Participants who completed at least 80% (19/24) of the prompts were included in a raffle to win a €100 (US $121.74) voucher and a cooking class. Before the second in-person session, EMA data were checked, and questions regarding implausible prompt entries (eg, 8 AM as bedtime) and food records (eg, missing meals) were collected. These questions were reviewed in the second in-person session to resolve any uncertainties. Usability of the APPetite-mobile-app, reactivity, and representativity of the EMA assessment were assessed via questionnaires in the second in-person session.

Participants received €40 (US $48.7) after the second in-person session and €10 (US $12.17) after completing the follow-up. In addition, individual feedback on diet and PA was provided after the follow-up, which consisted of a web-based 24-hour recall from home.

Statistical Methods

Descriptive statistics were used to assess feasibility measures. We investigated whether compliance differed among the 3 cohorts, the 3 days, and between male and female participants.

XSL•FO
RenderX

As compliance is not normally distributed, this is done using the following nonparametric tests: Kruskal-Wallis rank sum test, Friedman test, and Wilcoxon rank sum test. In addition, the Spearman rank correlation coefficient was calculated to investigate the association between compliance and age. The α level was set to .05. The analyses were performed using R 3.6.1 (R Core Team) with RStudio (RStudio, Public-benefit corporation).

Evaluation of Usability

Measures

Usability is assessed using the System Usability Scale (SUS; [54]), a commonly used questionnaire for the evaluation of websites or mobile apps. The questionnaire consists of 10 items. Each item represents a statement (eg, I thought the system was easy to use). Participants' agreement with the statement was rated on a 5-point scale. A total score between 0 and 100 was calculated. Higher numbers indicated better usability.

Sample

Data were collected within the APPetite study. However, SUS was subsequently added to this study (August 2019). Therefore, it is available only for a subsample of 84 participants (55 women and 29 men; 67 from LORA, 6 from PROUD, and 11 from BipoLife-A1). The mean age of the sample was 29.26 (SD 7.41) years, and the mean BMI was 24.82 (SD 5.26) kg/m^2.

Procedure

The SUS was completed during the second in-person session.

Statistical Methods

Total usability scores were calculated according to the study by Brooke [54] and presented through descriptive statistics (mean, SD, and range). We investigated whether usability was rated differently by the 3 cohorts using a one-way analysis of variance, as data are normally distributed and homogeneity of variance is given. An unpaired *t* test (two-tailed) was used to study gender differences, as assumptions of normal distribution and homogeneity of variance were met. The associations between usability and age (not normally distributed) as well as usability and compliance (not normally distributed) were investigated using Spearman rank correlations. The data were analyzed using R 3.6.1 with RStudio. The α level was set to .05.

Evaluation of Validity

Measures

The relative validity was assessed using a counterbalanced crossover design. Myfood24 Germany (Measure Your Food on One Day), a 24-hour recall, was chosen as the reference method. Myfood24 is a web-based, self-administered 24-hour dietary recall tool (refer to Koch et al [39] for details). It is based on two German nutritional databases (the German Food Code and Nutrient Data Base, Bundeslebensmittelschlüssel version 3.02, and the database LEBTAB of the Dortmund Nutritional and Anthropometric Longitudinally Designed study) and includes 11,501 food items. A comparison between habitual energy and macronutrient intake assessed through the APPetite-food record and 24-hour recall was drawn. Habitual intake was operationalized as the mean dietary intake of 3 days.

Furthermore, energy intake is compared with total energy expenditure (TEE) based on the assumption that energy intake equals TEE in weight-stable individuals [55]. TEE is estimated from nondominant wrist accelerometry according to White et al [56], which has been shown to be a precise approach to estimate TEE on population levels in free-living conditions when compared with TEE by doubly labeled water. The Euclidean norm minus one was extracted from the raw acceleration data using the DataAnalyzer from movisens (version 1.13.5; June 18, 2019) and inserted into the quadratic Euclidean norm minus one equation from White et al [56].

Sample

A total of 50 healthy participants from the LORA study (group 1: n=26; group 2: n=24) volunteered for the validation study. However, 6 participants from group 1 had to be excluded, as they did not complete all relevant parts within the predefined time schedule. Therefore, the evaluation of validity was based on data from 44 participants (33 women and 11 men)—20 from group 1 and 24 from group 2. This sample had a mean age of 28.64 (SD 8.13) years and a mean BMI of 23.8 (SD 3.62) kg/m^2. The groups did not significantly differ in terms of sex (group 1: 15 women and 5 men; group 2: 18 women and 6 men; X^2_1=0; *P*=.99), age (group 1: mean 30.15, SD 8.65 years; group 2: mean 27.38, SD 7.63 years; Mann-Whitney *U*=314; *P*=.08), and BMI (group 1: mean 23.99, SD 3.69 kg/m^2; group 2: mean 23.66, SD 3.63 kg/m^2; $t_{40.34}$=0.3 [unpaired; two-tailed]; *P*=.77).

Procedure

Participants from the LORA cohort who agreed to participate in the APPetite study were asked whether they wanted to also participate in the validation study: recording their food intake through a 24-hour recall on 3 additional days. Of the participants who agreed, 26 were assigned to group 1 and 24 to group 2, following a counterbalanced crossover design. Hence, participants in group 1 completed three 24-hour recalls exactly a week before the APPetite-food record was used. In group 2, participants completed three 24-hour recalls exactly a week after the APPetite-food record was used. The same weekdays, the week before or after, were assessed. Both groups received the same training to familiarize themselves with the 24-hour recall and the APPetite-food record. Participants received €30 (US $36.52) to participate in the validation study.

Statistical Methods

Habitual energy and macronutrient intake assessed through the APPetite-food record was compared with habitual dietary intake assessed through the 24-hour recall.

Habitual energy and macronutrient intake from the two methods were compared at the group level using two-tailed paired *t* tests (for normally distributed data including energy and carbohydrates) and Wilcoxon signed-rank tests (for skewed data including protein, fat, sugar, and fiber). Agreement between the two methods at the individual level was assessed using Bland-Altman analysis of the mean differences [57]. For this,

XSL•FO
RenderX

the difference between the two methods (y-axis) is plotted against the mean of the two methods (x-axis) for each participant. For reference, the mean difference between the two methods across all participants and the limits of agreement (LoA) estimated by the mean difference above and below 1.96 SD of the differences are shown in the plot. Thus, a systematic bias throughout the range of measurements can be identified. Acceptable LoA must be predefined. We predefined acceptable LoA for energy and macronutrient intake as 10% of the group mean across the two methods. Daily energy intake from the APPetite-food record (not normally distributed on days 2 and 3 of the EMA assessment) was compared with TEE (normally distributed) through a two-tailed paired *t* test for the first day and Wilcoxon signed-rank tests for the second and third days. Paired *t* tests (two-tailed) were calculated to compare mean TEE and habitual energy intake from the APPetite-food record and the 24-hour recall. The α level was set to .05. The analyses were performed using R 3.6.1 with RStudio.

Results

Feasibility

A total of 98.28% (3703/3768) of all scheduled prompts were delivered. The failure of prompt delivery was either due to technical problems or because the smartphone was switched off. Overall, 80.31% (3026/3768) of the prompts were answered completely. In total, 0.9% (34/3768) of prompts were registered as incomplete as a result of technical problems or extensive breaks during prompt completion. A total of 1.81% (68/3768) of prompts were dismissed, and 15.26% (575/3768) of prompts were ignored. The relatively large proportion of ignored prompts was, to some extent, a result of participants unintentionally leaving their smartphone at home or in another room. Furthermore, a number of participants reported that they had missed the first prompt or prompts of the day, as they were still sleeping.

Overall mean compliance (percentage of complete prompts within received prompts) was 81.73% (SD 21.65%). The compliance rate of 67.5% (106/157) of participants was above 80% (LORA: 94/136, 69.1%; PROUD: 4/7, 57%; and BipoLife-A1: 8/14, 57%). The mean compliance rate was 81.56% (SD 25.98%) on the first day, 83.28% (SD 23.55%) on the second day, and 79.97% (SD 25.8%) on the third day. The Friedman test showed no significant difference in compliance among the 3 days (X^2_2=3.6; *P*=.17), indicating no decline in motivation.

Compliance was highest in the LORA cohort and lowest in the PROUD cohort (see cohort means and SDs in Table 2). However, the Kruskal-Wallis rank sum test showed that compliance of the 3 cohorts did not differ significantly (X^2_2=0.7; *P*=.72).

Female participants had, on average, a compliance of 83.95% (SD 19.02%). The mean compliance for male participants was 77.83% (SD 25.34%). The Wilcoxon rank sum test found no gender difference in compliance (*P*=.16). No significant correlation was found between age and compliance (ρ=0.13; *P*=.12).

Participants responded to prompts after a mean of 189.32 seconds (SD 388.65). Responding to 70.54% (2157/3058) of all prompts was started within the first 60 seconds after the first prompt signal. The mean time needed to complete a single prompt was 122.63 seconds (SD 70.01). The prompt response latency and response duration for each of the 3 cohorts are shown in Table 3.

Table 2. Mean number and percentage of complete, incomplete, dismissed, and ignored prompts within received prompts for the total sample and each cohort.

Samples	Complete	Incomplete	Dismissed	Ignored
Total (N=157), mean (SD)				
Values	19.27 (5.32)	0.22 (0.44)	0.43 (1.07)	3.66 (4.79)
Percentage	81.73 (21.65)	0.93 (1.89)	1.82 (4.48)	15.52 (20.08)
LORA[a] (n=136), mean (SD)				
Values	19.43 (5.04)	0.23 (0.46)	0.4 (1.05)	3.5 (4.46)
Percentage	82.48 (20.41)	0.98 (1.95)	1.71 (4.4)	14.83 (18.7)
PROUD[b] (n=7), mean (SD)				
Values	17.14 (7.73)	0.43 (0.54)	0.86 (1.46)	5.57 (6.95)
Percentage	71.43 (32.22)	1.79 (2.23)	3.57 (6.09)	23.21 (28.95)
BipoLife-A1 (n=14), mean (SD)				
Values	18.86 (6.76)	0 (0)	0.5 (1.09)	4.29 (6.63)
Percentage	79.59 (27.46)	0 (0)	2.08 (4.55)	18.33 (27.8)

[a]LORA: Longitudinal Resilience Assessment.

[b]PROUD: Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder.

XSL•FO
RenderX

Table 3. Response latency and duration for a single prompt, reporting latency of the food record, and recording duration for the food record per day for the total sample and each cohort.

Variables	Total	LORA[a]	PROUD[b]	BipoLife-A1
Prompts (s), mean (SD)				
Latency	189.32 (388.65)	179.38 (375.82)	242.24 (447.85)	265.24 (469.8)
Duration	122.63 (70.01)	119.13 (64.21)	175.12 (144.31)	134.01 (64.07)
Food record (min), mean (SD)				
Latency	58.35 (127.52)	50.82 (115.8)	147.02 (197.85)	116.99 (190.16)
Duration	17.66 (8.66)	17.8 (8.57)	22.17 (10.86)	14.26 (7.56)

[a]LORA: Longitudinal Resilience Assessment.

[b]PROUD: Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder.

Dietary data of 8.8% (12/136) LORA, 14% (1/7) PROUD, and 14% (2/14) BipoLife-A1 participants had to be excluded, as the number of recorded meals or entered foods was evidently too low or entries were incomplete or implausible. In addition, 3 LORA participants had no food entry on 1 day. However, the remaining 2 days were recorded sufficiently well to be included. Included participants (n=142) recorded a total of 2969 eating and drinking events. In total, 3.03% (90/2969) entries were registered as incomplete, mainly due to technical problems. Participants entered on average 7.02 (SD 3.33) eating and drinking events per day (first day: mean 7.49, SD 3.14; range 2-17; no data available for 1 participant; second day: mean 7.08, SD 3.43; range 2-22; third day: mean 6.49, SD 3.37; range 2-17; no data available for 2 participants).

The mean latency from food intake to food recording was 58.35 (SD 127.52) minutes. Latency increased over the course of the 3 days (first day: mean 53.51, SD 72.01 min; second day: mean 69.5, SD 88.1 min; third day: 90.81, SD 116.12 min). The mean time to complete the food record of one day was 17.66 (SD 8.66) minutes. On the first day, participants took 21.01 (SD 9.68) minutes, on the second day they took 17.22 (SD 7.76) minutes, and on the third day they took 14.67 (SD 7.16) minutes. The cohort-specific food record latencies and recording durations are presented in Table 3.

The accelerometer records of 2 participants stopped during the second day. It is unknown if this was due to technical problems or because participants connected the sensor to a computer that instantly stopped the recording. In total, 11 participants did not wear the sensor on at least one day or stopped wearing it before the end of the 7-day assessment period. On average, participants (N=157 including the abovementioned) wore the sensor for 6 days 3 hours and 57 minutes (mean 8876.96, SD 1815.36 min; range 771-10,403). Hence, the mean compliance was 88.07% (SD 18.01%).

Usability

The SUS total score was 61.9 (SD 17.79; range 17.5-97.5) out of 100. The SUS score of the LORA cohort (n=67) was 61.23 (SD 16.8; range 17.5-95). The lowest usability with an SUS score of 60 (SD 24.08; range 32.5-92.5) was rated by the PROUD cohort (n=6). The highest SUS score was found for the Bipolife-A1 cohort (n=11), with a score of 67.05 (SD 20.97; range 22.5-97.5). However, the 3 cohorts did not differ in the ratings of usability according to a one-way analysis of variance ($F_{2,81}$=0.54; P=.59).

Female participants (mean 62.82, SD 17.36) scored usability on average marginally higher than male participants (mean 60.17, SD 18.77; t_{82}=0.65 [two-tailed]; P=.52). Age and usability were not significantly negatively correlated (ρ=−0.18; P=.10). Compliance and rated usability showed no significant correlation (ρ=0.13; P=.26).

Validity

Habitual intake of energy, protein, fat, carbohydrates, sugar, and fiber assessed through the APPetite-food record and the 24-hour recall are shown in Table 4. All nutritional intake was higher for the APPetite-food record. The difference between the two methods is significant for energy, protein, fat, and fiber intake (Table 4).

With regard to possible order effects, both groups (APPetite-food record first and 24-hour recall first) showed higher energy intake assessed through the APPetite-food record (group 1: 8494 kJ/2029 kcal; group 2: 9327 kJ/2228 kcal) compared with the 24-hour recall (group 1: 7881.23 kJ/1882 kcal; group 2: 8086.01 kJ/1931 kcal).

Agreement between the two methods at the individual level was investigated through Bland-Altman plots for energy and macronutrient intake. Mean energy difference between the APPetite-food record and the 24-hour recall was 994.18 kJ (95% CI 370.8-1617.6). A normal distribution of the difference was observed. Figure 3 shows the Bland-Altman plot of the habitual energy intake. The LoA are −3024.841 (95% CI −4104.6 to −1945.1) to 5013.2 (95% CI 3933.4-6093) and therefore larger than the predefined acceptable LoA of 849 kJ.

Bland-Altman analyses for protein, fat, carbohydrate, sugar, and fiber intake can be found in Multimedia Appendix 4 [57]. All LoA exceeded the predefined acceptable LoA.

XSL•FO
RenderX

Table 4. Mean habitual intake of energy and macronutrients from the APPetite-food record and the 24-hour recall; mean difference between the two methods; paired *t* tests (two-tailed) for normally distributed data including energy and carbohydrates; and Wilcoxon signed-rank tests for skewed data, including protein, fat, sugar, and fiber.

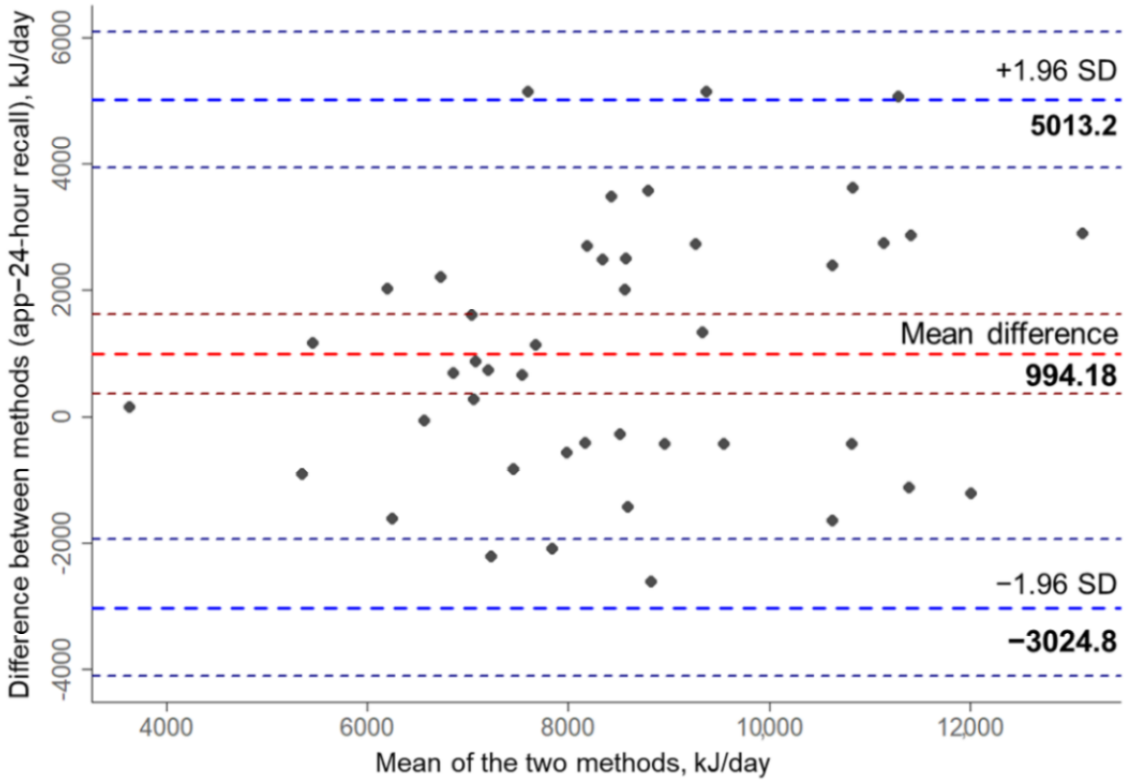
Dietary intake per day	APPetite-food record, mean (SD)	24-hour recall, mean (SD)	Mean difference (SD)	*t* test (*df*)	*P* value
Energy (kJ/day)	8987.11 (2404.65)	7992.93 (2002.61)	994.18 (2050.52)	3.21 (43)	.003[a]
Energy (kcal/day)	2146.42 (574.5)	1909.16 (478.8)	237.26 (489.94)	3.21 (43)	.003
Protein (g/day)	80.77 (27.6)	69.68 (24.72)	11.09 (22.63)	N/A[b]	.004[a]
Total fat (g/day)	92.25 (32.54)	76.95 (26.54)	15.3 (29.41)	N/A	.002[a]
Carbohydrate (g/day)	228.81 (63.12)	212.99 (52.65)	15.82 (64.03)	1.64 (43)	.11
Sugars (g/day)	81.21 (32.6)	76.97 (29.76)	4.24 (31.42)	N/A	.40
Fiber (g/day)	25.86 (9.22)	23.3 (8.5)	2.56 (7.43)	N/A	.04[c]

[a]*P*<.01.

[b]N/A: not applicable to Wilcoxon signed-rank test.

[c]*P*<.05.

Figure 3. Bland-Altman plot assessing agreement between habitual energy intake in kJ per day captured by the APPetite-food record and the 24-hour recall (red line: mean difference=app–24-hour recall; dark red lines: 95% CI of mean difference; blue lines: lower and upper limits of agreement; dark blue lines: 95% CI of lower and upper limits of agreement).



Energy intake from the APPetite-food record was significantly lower than the TEE estimated from accelerometry on the first day (t_{43}=5.33; *P*<.001; TEE mean 2425.4, SD 468.2; app mean 1897.53, SD 616.32), but did not significantly differ on days 2 and 3: day 2 (*P*=.051; TEE mean 2442.04, SD 447.5; app mean 2242.94, SD 769.78) and day 3 (*P*=.23; TEE mean 2435.6, SD 482.9; app mean 2317.77, SD 780.6). Mean TEE estimated from 7 days of accelerometry was 2417.8 kcal (SD 410) compared with the habitual energy intake of 2146.42 kcal (SD 574.5) from the APPetite-food record and 1909.16 kcal (SD 478.8) from the 24-hour recall. Paired *t* tests (two-tailed) showed that habitual energy intake was underestimated by both methods when compared with TEE: APPetite-food record (t_{43}=3.40; *P*=.002) and 24-hour recall (t_{43}=6.33; *P*<.001).

Discussion

Principal Findings

The APPetite-mobile-app was developed to capture complex dietary intake, objectively recorded PA, and related factors for studying microtemporal, within-person processes underlying

XSL•FO
RenderX

eating behavior and PA in daily life. This study evaluated the feasibility and usability of the EMA tool as well as the validity of the APPetite-food record. The APPetite-mobile-app demonstrated good feasibility. Compliance with responding to prompts and wearing the accelerometer was above 80%, and reasonable response times and latencies were found for the prompts as well as the food record. Usability was rated moderate, with a mean SUS score of 61.9. Large differences in energy and macronutrient intake assessed with the APPetite-food record versus the 24-hour recall were found at the group and individual levels, whereby the APPetite-food record captured higher dietary intakes. Energy intake was assessed fairly accurately by the APPetite-food record on the group level on 2 of 3 days when compared with TEE. The comparison of habitual energy intake to mean TEE showed that the 24-hour recall underestimated energy intake to a larger degree than the APPetite-food record. These results indicate that the discrepancies between the two dietary assessment methods do not imply a lack of validity of the APPetite-food record; rather, they indicate a more accurate dietary assessment compared with the 24-hour recall and therefore provide the first evidence that the APPetite-food record is a valid tool for capturing complex dietary intake.

Comparison With Previous Work

The good feasibility of EMA tools is crucial to ensure unbiased data collection and prevent systematic missing data. Compliance rates are an important indicator of feasibility. Although there is no official criterion indicating good compliance, Stone and Shiffman [58] proposed compliance rates above 80% to be acceptable. However, they emphasized the arbitrariness of this criterion and the need to define acceptable compliance ranges for each study individually, especially when noncompliance may be systematic and not random. The mean prompt compliance in our study was above 80% and can therefore generally be rated as good. These good compliance rates may be partly due to the notifications participants received when falling below the 80% threshold and the incentive to be included in raffles when reaching a compliance rate of 80% or above.

Furthermore, the results demonstrate that prompt compliance did not decrease over the course of 3 days. In other EMA studies that assessed more than 3 days, response rates declined substantially (eg, 40% from 63% on day 1 to 23% on day 7), even with only 4 prompts per day [36]. As studying microtemporal processes requires an illustration of a day in high resolution, it is more important to focus on a larger number of completed prompts per day compared with a large number of EMA assessment days. On the basis of our constant compliance rate over 3 days, the length of the EMA assessment seems feasible, and no decline in motivation was evident.

We found marginally lower prompt compliance rates in the clinical cohorts than in the healthy cohort. In a study by Porras-Segovia et al [59] comparing EMA compliance rates from suicidal patients and student controls, lower compliance was found for the clinical sample. These findings were consistent with our results. However, Porras-Segovia et al [59] found a significant difference between patients and healthy controls, which was not the case in our study. These results suggest that the prompt schedule of the APPetite-mobile-app is equally well suited for healthy individuals and patients with a mental disorder.

A mean of 2.04 (SD 1.17) minutes was needed to complete 1 prompt. In accordance with the high prompt compliance rate, the response duration of the prompts can be considered feasible. Responding to a prompt was initiated on average 3.16 (SD 6.48) minutes after the first prompt signal. Short prompt latencies are essential to guarantee the momentary nature of the response and should therefore be taken into account thoroughly. However, most studies have not reported prompt latencies [60,61]. Some studies have predefined response windows. This ensures the momentary nature of the response but can cause lower compliance rates, for example, 69% in a study with an 8-minute response window [62]. We chose to allow a longer response period and prompt postponement of up to 25 minutes to reduce participants' burden and maintain high compliance. Nevertheless, participants were instructed to respond to EMA prompts instantly, if possible. Considering that we allowed responses up to 30 minutes after the first prompt signal, the mean latency of just over 3 minutes is short and underlines the feasibility of the prompts.

Compliance with the food record cannot be directly determined, as it is not possible to differentiate between someone not recording a food item because of noncompliance or because of not actually consuming it. However, other quality measures could also be used. The time spent reporting daily dietary intake or the number of recorded eating and drinking occasions per day can be used for quality checks. On average, participants needed 17.6 minutes to complete the food record of 1 day. Other technology-based tools for assessing dietary intake show similar times to complete, ranging between 13 and 45 minutes [20]. Participants entered on average 7 eating and drinking events per day. This number is in line with a previous study that found a mean of 20.7 eating and drinking occasions per individual over a 3-day period [29].

The APPetite-food record was developed to record food intake in real time or near real time. Therefore, it is important to consider the amount of time between food intake and recording. Foods and drinks were recorded on average 58.35 minutes after intake. This shows that participants did not wait until the evening to record all eating and drinking occasions for 1 day. Hence, food intake was recorded in real time or near real time.

The food recording behavior of our participants suggests that the APPetite-food record is feasible. However, we noticed that the participants' motivation was crucial for successfully capturing sufficient and accurate dietary data. Training is needed to ensure that participants understand the importance of food recording in real time or near real time. Furthermore, participants reported that receiving detailed dietary feedback at the end of the study increased their motivation to enter food intake accurately.

As expected from previous studies [63], a high compliance rate (88.07%) for the accelerometer worn on the wrist were found. All measures of feasibility regarding prompts, the APPetite-food record, and the accelerometer indicate that the APPetite-mobile-app is a feasible EMA tool.

XSL•FO
RenderX

In addition to good feasibility, usability is an important criterion that should be considered when developing new EMA tools. The usability of the APPetite-mobile-app was rated as moderate, with an SUS score of 61.9 out of 100. In a previous study, the usability of the top 7 iPhone operating system and Android diet-tracking apps was assessed [22]. The usability of 2 apps was rated even lower than the APPetite-mobile-app (Lose It!=59.2; MyDietCoach=46.7). However, a comparison is difficult as these tools focus purely on dietary assessment. The SUS score of the APPetite-mobile-app was rated on the basis of both dietary assessment and EMA prompts. The relatively low usability of our tool can be explained by the fact that it is a scientific device and was therefore developed independently without professional app developers. High costs are involved in the professional development of an app. For this reason, we chose the platform movisensXS to independently develop the app. Although movisensXS has many configuration options, it still has its limitations. For example, a search function within the food record cannot be implemented. The app was developed for scientific purposes only and not for consumer use. However, usability challenges have been reported even for commercial tools; for instance, only 20% of participants would continue to use MyFitnessPal after study participation [64]. Even though the usability of the APPetite-mobile-app was rated relatively low, no negative effect on feasibility, including compliance, was evident. Therefore, an improvement in usability is desirable but not essential for its use in scientific research.

A food record was incorporated into the APPetite-mobile-app to capture complex dietary intake in real time or near real time. An evaluation of validity was needed to test whether the APPetite-food record accurately assessed dietary intake. Hence, the APPetite-food record was compared with a 24-hour recall and TEE estimated from nondominant wrist accelerometry.

With regard to relative validity, low agreement between habitual dietary intake measured by the APPetite-food record and the 24-hour recall was found at both the group and individual levels. At the group level, energy, protein, fat, and fiber intake from the APPetite-food record was significantly higher than the 24-hour recall. Wide LoA, which exceeded the predefined acceptable LoA, were found for energy, protein, fat, carbohydrate, sugar, and fiber intake at the individual level. One could argue that these discrepancies indicate a lack of validity in the APPetite-food record. However, even though 24-hour recalls are frequently used as the established reference method when assessing relative validity, the true intake remains unknown [65]. Therefore, possible reasons for this discrepancy must be taken into account for both methods. Most validation studies that compared an EMA dietary assessment tool with a 24-hour recall found lower values for energy intake as well as intake of some macronutrients assessed through the EMA tool on the descriptive or even statistical level (eg, [64] for energy [statistical], proteins [statistical], fat [statistical], carbohydrates [statistical], and sugar [statistical]; [65] for energy, protein, fat, and carbohydrates; [66] for energy and fat [statistical], not for proteins and carbohydrates; [67] for energy, fat, and carbohydrates, not for proteins; [68] for energy, protein, sugar, and fat, not for carbohydrates and fiber, no statistical hypothesis test reported; [69] for energy, fat [statistical], carbohydrates, and fiber, not for proteins). This was not the case in our study. Habitual energy, protein, fat, carbohydrate, sugar, and fiber intake was higher when assessed with the APPetite-food record on the descriptive or even statistical level, indicating a lower degree of underreporting. This leads to the conclusion that the APPetite-food record could be a more precise dietary assessment method than the 24-hour recall.

This interpretation is underlined by the comparison of energy intake and TEE estimated from accelerometry. Energy intake from the APPetite-food record was not significantly different from TEE on 2 of 3 days, indicating that the APPetite-food record assesses energy intake fairly accurately at the group level. However, the comparison with the mean TEE showed that both methods underestimated habitual energy intake. In this context, it must be mentioned that over one-third of the participants in the validation study (17/44, 39%) indicated that they are currently trying to lose weight. Therefore, the discrepancy between the TEE and the reported energy intake could, to some extent, be due to diet and weight loss. However, the 24-hour recall underestimated habitual energy intake to a greater extent. Inaccurate estimates of energy intake captured by 24-hour recalls have been reported in previous studies [70]. A reason for the improved reporting accuracy in dietary assessments in real time or near real time compared with retrospective assessments could be the minimized retention interval [71,72]. Memory effects can cause bias in retrospective dietary assessments, as the demand for memory increases simultaneously with the retention interval. Memory lapses can cause two types of errors in the context of 24-hour dietary recalls: the failure to recall foods actually consumed (errors of omission) and the reporting of foods that were actually not consumed during the recalled day (errors of intrusion) [73]. Furthermore, incorrect estimations of portion sizes have been reported to constitute the largest measurement error in 24-hour recalls [73]. This error is closely related to memory bias, as consumed amounts must not only be accurately estimated but also be correctly remembered [74]. Food records in real time or near real time can minimize memory errors [74]. In a recent study, 65% of participants reported that remembering meal items and portion sizes was easier in a progressive assessment than in a traditional retrospective 24-hour recall [75]. Nonetheless, food records in real time or near real time are also affected by potential bias, which was also shown in our study as underestimation of food intake became evident. In particular, the change in dietary intake as a result of recording it has to be taken into account. Participants may choose not to eat complex meals or eat less to avoid extensive and time-consuming records [74]. Furthermore, keeping a record of food intake in daily life can be burdensome. Participants may not be able to record everything eaten due to other commitments. However, our results suggest that the impact of reactivity and high burden on the APPetite-food record might be smaller than the effect of memory loss on 24-hour recalls.

The results of the evaluation of validity indicate that the APPetite-food record might assess dietary intake more accurately than the 24-hour recall and capture daily energy intake fairly accurately at the group level. Nevertheless, both dietary assessment methods seem to underestimate habitual energy intake.

XSL•FO
RenderX

Strengths and Limitations

A common validation approach is the assessment of relative validity, which compares a novel tool to an established dietary assessment method. However, most of the available validation studies show methodological issues as they assess relative validity on overlapping days [64-68] and do not use a counterbalanced crossover design [69]. Assessing overlapping days can lead to an overestimation of agreement between two self-report methods, as recording dietary intake actively throughout the day may improve memory for completing the 24-hour dietary recall of the same day. Empirical evidence for the overestimated agreement has been found, detecting improved accuracy of 24-hour recalls of days when diet was tracked throughout [65]. A further problem becomes apparent in studies that do not assess overlapping days. When two methods are used one after another, order effects can bias the assessment. However, most studies did not control for possible order effects [69]. We were able to counteract these methodological issues by choosing a counterbalanced crossover study design that assessed no overlapping days. A counterbalanced crossover design is crucial for controlling learning, boredom, and other unwanted order effects. We understand this to be the most significant strength of our validation study.

One limitation of our validation study is due to the fact that dietary intake varies from day to day. Bland and Altman call this case "method where true value varies" [76]. When the true value varies, measurements of two methods have to be taken at the same time point to obtain an accurate estimate of agreement [77]. In the context of dietary assessment methods that would translate to assessing food intake using two methods on the same day. However, because an inflated agreement when assessing overlapping days is likely to occur [26], as mentioned earlier, this does not represent a suitable approach. Therefore, we were not able to compare dietary intake on a day level (eg, Thursday compared with Thursday the week before or after) and chose to compare habitual dietary intake instead. However, when comparing habitual dietary intake, two aspects must be considered: (1) the target of interest of the APPetite-mobile-app is not regular or habitual food intake but rather microtemporal dynamics of food intake in daily life. Using habitual intake as the measure of comparison sets aside this fact and might therefore not be the most appropriate measure for the evaluation of validity. (2) Day-to-day variability in dietary intake represents a problem when assessing habitual intake. It could be argued that capturing 3 days to operationalize habitual intake is not sufficient to obtain an accurate estimate.

Many studies that use Bland-Altman agreement analyses to evaluate the validity of food records in real time or near real time have inaccuracies. To the best of our knowledge, our study is the only one that has a predefined acceptable LoA. These pre-established limits are necessary to avoid misleading interpretations. A consensus on the acceptable LoA for dietary intake is desirable. This will improve the comparability of the results from studies assessing relative validity. Furthermore, the use of established but biased dietary assessment methods, such as 24-hour recalls, to study relative validity should be questioned critically. New approaches to evaluate the validity of food records in real time or near real time are needed.

Our findings are limited because of the lack of control for possible weight changes during study participation. The comparison of TEE and energy intake is based on the assumption that energy expenditure is equal to energy intake. However, this assumption is valid only for weight-stable individuals.

Two further limitations concern the APPetite-food record itself: (1) nutritional values are generated manually, which is time-consuming and can be error-prone. Automated generation is preferable. (2) The APPetite-food record relies on self-reports of dietary intake. Self-reports are subjective and therefore more likely to be biased. To add a more objective component to the dietary assessment, photos of the foods and drinks consumed could be taken in addition to self-reports.

The strength of our assessment of feasibility and usability is that the sample of healthy participants was enriched with data from patients suffering from a mental disorder. Therefore, it was possible to show that the APPetite-mobile-app is equally feasible and usable in this population. This finding is particularly important as diet and PA play an important role in mental health. This opens up the possibility of studying microtemporal, within-person processes of diet, PA, and related factors in psychiatric patients, which is crucial for the understanding of the link among diet, PA, and mental health. However, the unequal sample sizes of the 3 cohorts limit the results. This is of concern in the context of cohort comparisons, as well as the interpretation of the means of the total sample. Furthermore, a selection bias could be present, as the participants were exclusively recruited from 3 existing study cohorts.

Recommendations for Future Studies

The development of novel EMA tools for assessing microtemporal processes of diet, PA, and related factors is required. Studies comparing these new EMA tools are needed to establish empirical evidence on which assessment approaches are most effective in the study of microtemporal processes. Future EMA studies should consider that participants' motivation is the key to success, especially when complex dietary intake is assessed. Therefore, participants' burden needs to be kept minimal, and incentives for prompt responding and food recording, such as dietary feedback and raffle inclusions, are essential.

New technologies and wearable sensors are a promising advancement in the area of dietary assessment in naturalistic settings, as they can passively detect eating behavior [78]. They can be used for longer assessment periods because they require minimal user interaction. These sensors will improve the validity of self-reported dietary assessments to a great extent. We believe they will soon be of tremendous relevance, especially for the assessment of microtemporal processes of diet in daily life.

Conclusions

The APPetite-mobile-app is a promising tool for studying microtemporal, within-person processes of diet, PA, and related factors in real time or near real time and is, to the best of our knowledge, the first of its kind. First evidence supports that the APPetite-mobile-app is feasible and the APPetite-food record is a valid tool for capturing complex dietary intake. We hope

XSL•FO
RenderX

this motivates other researchers to use EMA to capture complex dietary intake, PA, and associated factors in daily life, and it initiates a discussion about feasible, usable, and valid methods to assess these dynamics. Assessment strategies need to be developed, shared, and discussed to advance the research field. A solid empirical foundation regarding within-person, microtemporal associations of diet, PA, and associated factors is needed for the development of personalized lifestyle modification interventions, such as intensively adaptive interventions or just-in-time adaptive interventions [10].

Acknowledgments

This work was supported by the European Union's Horizon 2020 Research and Innovation Program under grant agreement number 728018. The funding source had no involvement in the study design, data collection, interpretation of the findings, or writing of this manuscript.

Conflicts of Interest

UEP reports consultancy for Boehringer-Ingelheim. All other authors declare no conflicts of interest.

Multimedia Appendix 1
APPetite-food record and nutritional value generation.
[DOCX File , 111 KB - jmir_v23i7e25850_app1.docx]

Multimedia Appendix 2
Reasons for food intake.
[DOCX File , 33 KB - jmir_v23i7e25850_app2.docx]

Multimedia Appendix 3
Prompt measures and items.
[DOCX File , 39 KB - jmir_v23i7e25850_app3.docx]

Multimedia Appendix 4
Bland-Altman analysis of protein, fat, carbohydrate, sugar, and fiber intake.
[DOCX File , 415 KB - jmir_v23i7e25850_app4.docx]

References

1. World Health Organization. Obesity and Overweight.: World Health Organization; 2020. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-09-25]
2. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
3. Adan RA, van der Beek EM, Buitelaar JK, Cryan JF, Hebebrand J, Higgs S, et al. Nutritional psychiatry: towards improving mental health by what you eat. Eur Neuropsychopharmacol 2019 Dec;29(12):1321-1332 [FREE Full text] [doi: 10.1016/j.euroneuro.2019.10.011] [Medline: 31735529]
4. Kelly SJ, Daniel M, Grande ED, Taylor A. Mental ill-health across the continuum of body mass index. BMC Public Health 2011 Oct 05;11:765 [FREE Full text] [doi: 10.1186/1471-2458-11-765] [Medline: 21975214]
5. Rajan TM, Menon V. Psychiatric disorders and obesity: a review of association studies. J Postgrad Med 2017;63(3):182-190 [FREE Full text] [doi: 10.4103/jpgm.JPGM_712_16] [Medline: 28695871]
6. Milaneschi Y, Simmons WK, van Rossum EF, Penninx BW. Depression and obesity: evidence of shared biological mechanisms. Mol Psychiatry 2019 Jan;24(1):18-33. [doi: 10.1038/s41380-018-0017-5] [Medline: 29453413]
7. Sleddens EF, Kroeze W, Kohl LF, Bolten LM, Velema E, Kaspers P, et al. Correlates of dietary behavior in adults: an umbrella review. Nutr Rev 2015 Aug;73(8):477-499 [FREE Full text] [doi: 10.1093/nutrit/nuv007] [Medline: 26106126]
8. Sleddens EF, Kroeze W, Kohl LF, Bolten LM, Velema E, Kaspers PJ, et al. Determinants of dietary behavior among youth: an umbrella review. Int J Behav Nutr Phys Act 2015 Feb 01;12:7 [FREE Full text] [doi: 10.1186/s12966-015-0164-x] [Medline: 25638322]
9. Wood W, Neal DT. Healthy through habit: Interventions for initiating & maintaining health behavior change. Behav Sci Policy 2016;2(1):71-83 [FREE Full text] [doi: 10.1353/bsp.2016.0008]
10. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. J Am Med Assoc 2018 Aug 21;320(7):639-640. [doi: 10.1001/jama.2018.6621] [Medline: 29852046]

XSL•FO
RenderX

11. Hummel E, Hoffmann I. Complexity of nutritional behavior: capturing and depicting its interrelated factors in a cause-effect model. Ecol Food Nutr 2016;55(3):241-257. [doi: 10.1080/03670244.2015.1129325] [Medline: 26828451]
12. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
13. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
14. Bleich SN, Jones-Smith J, Wolfson JA, Zhu X, Story M. The complex relationship between diet and health. Health Aff (Millwood) 2015 Nov;34(11):1813-1820. [doi: 10.1377/hlthaff.2015.0606] [Medline: 26526238]
15. Moore MS. Interactions between physical activity and diet in the regulation of body weight. Proc Nutr Soc 2000 May;59(2):193-198. [doi: 10.1017/s0029665100000215] [Medline: 10946787]
16. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003;4(1):35-52. [doi: 10.1023/A:1023657221954]
17. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
18. Maugeri A, Barchitta M. A systematic review of ecological momentary assessment of diet: implications and perspectives for nutritional epidemiology. Nutrients 2019 Nov 07;11(11):2696 [FREE Full text] [doi: 10.3390/nu11112696] [Medline: 31703374]
19. Mason TB, Do B, Wang S, Dunton GF. Ecological momentary assessment of eating and dietary intake behaviors in children and adolescents: a systematic review of the literature. Appetite 2020 Jan 01;144:104465. [doi: 10.1016/j.appet.2019.104465] [Medline: 31541670]
20. Eldridge AL, Piernas C, Illner A, Gibney MJ, Gurinović MA, de Vries JH, et al. Evaluation of new technology-based tools for dietary intake assessment-an ILSI Europe dietary intake and exposure task force evaluation. Nutrients 2018 Dec 28;11(1):55 [FREE Full text] [doi: 10.3390/nu11010055] [Medline: 30597864]
21. Pendergast FJ, Leech RM, McNaughton SA. Novel online or mobile methods to assess eating patterns. Curr Nutr Rep 2017;6(3):212-227 [FREE Full text] [doi: 10.1007/s13668-017-0211-0] [Medline: 28944099]
22. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
23. Jeffers AJ, Mason TB, Benotsch EG. Psychological eating factors, affect, and ecological momentary assessed diet quality. Eat Weight Disord 2020 Oct;25(5):1151-1159. [doi: 10.1007/s40519-019-00743-3] [Medline: 31388844]
24. Roy PG, Jones KK, Martyn-Nemeth P, Zenk SN. Contextual correlates of energy-dense snack food and sweetened beverage intake across the day in African American women: an application of ecological momentary assessment. Appetite 2019 Jan 01;132:73-81. [doi: 10.1016/j.appet.2018.09.018] [Medline: 30261234]
25. Powell DJ, McMinn D, Allan JL. Does real time variability in inhibitory control drive snacking behavior? An intensive longitudinal study. Health Psychol 2017 Dec;36(4):356-364. [doi: 10.1037/hea0000471] [Medline: 28192005]
26. Richard A, Meule A, Reichenberger J, Blechert J. Food cravings in everyday life: an EMA study on snack-related thoughts, cravings, and consumption. Appetite 2017 Jun 01;113:215-223. [doi: 10.1016/j.appet.2017.02.037] [Medline: 28249745]
27. Zenk SN, Horoi I, McDonald A, Corte C, Riley B, Odoms-Young AM. Ecological momentary assessment of environmental and personal factors and snack food intake in African American women. Appetite 2014 Dec;83:333-341. [doi: 10.1016/j.appet.2014.09.008] [Medline: 25239402]
28. Allan J, McMinn D, Powell D. Tracking snacking in real time: time to look at individualised patterns of behaviour. Nutr Health 2019 Sep;25(3):179-184. [doi: 10.1177/0260106019866099] [Medline: 31347450]
29. Franja S, Wahl DR, Elliston KG, Ferguson SG. Comfort eating: an observational study of affect in the hours immediately before, and after, snacking. Br J Health Psychol 2021 Jan 04:12505. [doi: 10.1111/bjhp.12505] [Medline: 33398921]
30. Ashurst J, van Woerden I, Dunton G, Todd M, Ohri-Vachaspati P, Swan P, et al. The association among emotions and food choices in first-year college students using mobile-ecological momentary assessments. BMC Public Health 2018 May 02;18(1):573 [FREE Full text] [doi: 10.1186/s12889-018-5447-0] [Medline: 29716572]
31. Grenard JL, Stacy AW, Shiffman S, Baraldi AN, MacKinnon DP, Lockhart G, et al. Sweetened drink and snacking cues in adolescents: a study using ecological momentary assessment. Appetite 2013 Aug;67:61-73 [FREE Full text] [doi: 10.1016/j.appet.2013.03.016] [Medline: 23583312]
32. Bruening M, Ohri-Vachaspati P, Brewis A, Laska M, Todd M, Hruschka D, et al. Longitudinal social networks impacts on weight and weight-related behaviors assessed using mobile-based ecological momentary assessments: Study Protocols for the SPARC study. BMC Public Health 2016 Dec 30;16:901 [FREE Full text] [doi: 10.1186/s12889-016-3536-5] [Medline: 27576358]
33. Wahl DR, Villinger K, König LM, Ziesemer K, Schupp HT, Renner B. Healthy food choices are happy food choices: Evidence from a real life sample using smartphone based assessments. Sci Rep 2017 Dec 06;7(1):17069. [doi: 10.1038/s41598-017-17262-9] [Medline: 29213109]

XSL•FO
RenderX

34. Seto E, Hua J, Wu L, Shia V, Eom S, Wang M, et al. Models of individual dietary behavior based on smartphone data: the influence of routine, physical activity, emotion, and food environment. PLoS One 2016;11(4):e0153085 [FREE Full text] [doi: 10.1371/journal.pone.0153085] [Medline: 27049852]
35. Schüz B, Bower J, Ferguson SG. Stimulus control and affect in dietary behaviours. An intensive longitudinal study. Appetite 2015 Apr;87:310-317. [doi: 10.1016/j.appet.2015.01.002] [Medline: 25579222]
36. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
37. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
38. Adamo KB, Prince SA, Tricco AC, Connor-Gorber S, Tremblay M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. Int J Pediatr Obes 2009;4(1):2-27. [doi: 10.1080/17477160802315010] [Medline: 18720173]
39. Koch SA, Conrad J, Hierath L, Hancock N, Beer S, Cade JE, et al. Adaptation and evaluation of Myfood24-Germany: a web-based self-administered 24-h dietary recall for the German adult population. Nutrients 2020 Jan 06;12(1):160 [FREE Full text] [doi: 10.3390/nu12010160] [Medline: 31935885]
40. Cleobury L, Tapper K. Reasons for eating 'unhealthy' snacks in overweight and obese males and females. J Hum Nutr Diet 2014 Aug;27(4):333-341. [doi: 10.1111/jhn.12169] [Medline: 24134077]
41. Verhoeven AA, Adriaanse MA, de Vet E, Fennis BM, de Ridder DT. It's my party and I eat if I want to. Reasons for unhealthy snacking. Appetite 2015 Jan;84:20-27. [doi: 10.1016/j.appet.2014.09.013] [Medline: 25261101]
42. Schüz B, Papadakis T, Ferguson SG. Situation-specific social norms as mediators of social influence on snacking. Health Psychol 2018 Feb;37(2):153-159. [doi: 10.1037/hea0000568] [Medline: 29154607]
43. Renner B, Sproesser G, Strohbach S, Schupp HT. Why we eat what we eat. The Eating Motivation Survey (TEMS). Appetite 2012 Aug;59(1):117-128. [doi: 10.1016/j.appet.2012.04.004] [Medline: 22521515]
44. Evers C, Adriaanse M, de Ridder DT, de Witt Huberts JC. Good mood food. Positive emotion as a neglected trigger for food intake. Appetite 2013 Sep;68:1-7. [doi: 10.1016/j.appet.2013.04.007] [Medline: 23602962]
45. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015 Dec;34(12):1145-1153. [doi: 10.1037/hea0000223] [Medline: 26053885]
46. Wilhelm P, Schoebi D. Assessing mood in daily life: structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. Eur J Psychol Assess 2007 Jan;23(4):258-267. [doi: 10.1027/1015-5759.23.4.258]
47. Reichenberger J, Kuppens P, Liedlgruber M, Wilhelm FH, Tiefengrabner M, Ginzinger S, et al. No haste, more taste: an EMA study of the effects of stress, negative and positive emotions on eating behavior. Biol Psychol 2018 Jan;131:54-62. [doi: 10.1016/j.biopsycho.2016.09.002] [Medline: 27654506]
48. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
49. Tomko RL, Solhan MB, Carpenter RW, Brown WC, Jahng S, Wood PK, et al. Measuring impulsivity in daily life: the momentary impulsivity scale. Psychol Assess 2014 Jun;26(2):339-349 [FREE Full text] [doi: 10.1037/a0035083] [Medline: 24274047]
50. Chmitorz A, Neumann RJ, Kollmann B, Ahrens KF, Öhlschläger S, Goldbach N, et al. Longitudinal determination of resilience in humans to identify mechanisms of resilience to modern-life stressors: the longitudinal resilience assessment (LORA) study. Eur Arch Psychiatry Clin Neurosci 2020 Jul 18:A. [doi: 10.1007/s00406-020-01159-2] [Medline: 32683526]
51. Mayer JS, Hees K, Medda J, Grimm O, Asherson P, Bellina M, et al. Bright light therapy versus physical exercise to prevent co-morbid depression and obesity in adolescents and young adults with attention-deficit / hyperactivity disorder: study protocol for a randomized controlled trial. Trials 2018 Feb 26;19(1):140. [doi: 10.1186/s13063-017-2426-1] [Medline: 29482662]
52. Pfennig A, Leopold K, Martini J, Boehme A, Lambert M, Stamm T, et al. Improving early recognition and intervention in people at increased risk for the development of bipolar disorder: study protocol of a prospective-longitudinal, naturalistic cohort study (Early-BipoLife). Int J Bipolar Disord 2020 Jul 01;8(1):22 [FREE Full text] [doi: 10.1186/s40345-020-00183-4] [Medline: 32607662]
53. Ritter PS, Bermpohl F, Gruber O, Hautzinger M, Jansen A, Juckel G, et al. Aims and structure of the German Research Consortium BipoLife for the study of bipolar disorder. Int J Bipolar Disord 2016 Dec;4(1):26 [FREE Full text] [doi: 10.1186/s40345-016-0066-0] [Medline: 27873290]
54. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland AL, Weerdmeester BA, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996:1-6.
55. Trabulsi J, Schoeller DA. Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. Am J Physiol Endocrinol Metab 2001 Nov;281(5):891-899. [doi: 10.1152/ajpendo.2001.281.5.E891] [Medline: 11595643]

XSL•FO
RenderX

56. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. Int J Obes (Lond) 2019 Nov;43(11):2333-2342 [FREE Full text] [doi: 10.1038/s41366-019-0352-x] [Medline: 30940917]
57. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
58. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
59. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 01;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
60. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
61. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based Ecological Momentary Assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017 Nov 01;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
62. Rusby JC, Westling E, Crowley R, Light JM. Psychosocial correlates of physical and sedentary activities of early adolescent youth. Health Educ Behav 2014 Feb;41(1):42-51 [FREE Full text] [doi: 10.1177/1090198113485753] [Medline: 23640122]
63. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014 Jul;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
64. Chen J, Berkman W, Bardouh M, Ng CY, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]
65. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
66. Bucher DT, Carrard I, Farina E, Danuser B, Kruseman M. Development and evaluation of e-CA, an electronic mobile-based food record. Nutrients 2017 Jan 18;9(1):76 [FREE Full text] [doi: 10.3390/nu9010076] [Medline: 28106767]
67. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 09;17(1):5 [FREE Full text] [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]
68. Rangan AM, O'Connor S, Giannelli V, Yap ML, Tang LM, Roy R, et al. Electronic Dietary Intake Assessment (e-DIA): comparison of a mobile phone digital entry app for dietary data collection with 24-hour dietary recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98 [FREE Full text] [doi: 10.2196/mhealth.4613] [Medline: 26508282]
69. Ashman AM, Collins CE, Brown LJ, Rae KM, Rollo ME. Validation of a smartphone image-based dietary assessment method for pregnant women. Nutrients 2017 Jan 18;9(1):73 [FREE Full text] [doi: 10.3390/nu9010073] [Medline: 28106758]
70. Lopes TS, Luiz RR, Hoffman DJ, Ferriolli E, Pfrimer K, Moura AS, et al. Misreport of energy intake assessed with food records and 24-h recalls compared with total energy expenditure estimated with DLW. Eur J Clin Nutr 2016 Nov;70(11):1259-1264. [doi: 10.1038/ejcn.2016.85] [Medline: 27273069]
71. Baxter SD, Thompson WO, Davis HC, Johnson MH. Impact of gender, ethnicity, meal component, and time interval between eating and reporting on accuracy of fourth-graders' self-reports of school lunch. J Am Diet Assoc 1997 Nov;97(11):1293-1298. [doi: 10.1016/S0002-8223(97)00309-X] [Medline: 9366868]
72. Baxter SD, Hardin JW, Guinn CH, Royer JA, Mackelprang AJ, Smith AF. Fourth-grade children's dietary recall accuracy is influenced by retention interval (target period and interview time). J Am Diet Assoc 2009 May;109(5):846-856 [FREE Full text] [doi: 10.1016/j.jada.2009.02.015] [Medline: 19394471]
73. Gibson RS, Charrondiere UR, Bell W. Measurement errors in dietary assessment using self-reported 24-hour recalls in low-income countries and strategies for their prevention. Adv Nutr 2017 Nov;8(6):980-991. [doi: 10.3945/an.117.016980] [Medline: 29141979]
74. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51 [FREE Full text] [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
75. Osadchiy T, Poliakov I, Olivier P, Rowland M, Foster E. Progressive 24-hour recall: usability study of short retention intervals in web-based dietary assessment surveys. J Med Internet Res 2020 Feb 03;22(2):e13266 [FREE Full text] [doi: 10.2196/13266] [Medline: 32012055]
76. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat 2007;17(4):571-582. [doi: 10.1080/10543400701329422] [Medline: 17613642]
77. Olofsen E, Dahan A, Borsboom G, Drummond G. Improvements in the application and reporting of advanced Bland-Altman methods of comparison. J Clin Monit Comput 2015 Feb;29(1):127-139. [doi: 10.1007/s10877-014-9577-3] [Medline: 24806333]

XSL•FO
RenderX

78. Bell BM, Alam R, Alshurafa N, Thomaz E, Mondol AS, de la Haye K, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. NPJ Digit Med 2020;3:38 [FREE Full text] [doi: 10.1038/s41746-020-0246-2] [Medline: 32195373]

Abbreviations

EMA: ecological momentary assessment
LoA: limits of agreement
LORA: Longitudinal Resilience Assessment
PA: physical activity
PROUD: Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder
SUS: System Usability Scale
TEE: total energy expenditure

Edited by R Kukafka; submitted 22.11.20; peer-reviewed by J Reichenberger, I Gabashvili; comments to author 01.01.21; revised version received 12.01.21; accepted 04.05.21; published 05.07.21.

Please cite as:
Ruf A, Koch ED, Ebner-Priemer U, Knopf M, Reif A, Matura S
Studying Microtemporal, Within-Person Processes of Diet, Physical Activity, and Related Factors Using the APPetite-Mobile-App: Feasibility, Usability, and Validation Study
J Med Internet Res 2021;23(7):e25850
URL: https://www.jmir.org/2021/7/e25850
doi:10.2196/25850
PMID:34342268

©Alea Ruf, Elena Doris Koch, Ulrich Ebner-Priemer, Monika Knopf, Andreas Reif, Silke Matura. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Australian Children's Exposure to, and Engagement With, Web-Based Marketing of Food and Drink Brands: Cross-sectional Observational Study

Bridget Kelly[1], BSc, MPH, PhD; Rebecca Bosward[1], BPH; Becky Freeman[2], BSc, MSc, PhD

[1]Early Start, School of Health & Society, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia
[2]Prevention Research Collaboration, School of Public Heath, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

Corresponding Author:
Bridget Kelly, BSc, MPH, PhD
Early Start, School of Health & Society
Faculty of the Arts, Social Sciences and Humanities
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 2 4221 3893
Email: bkelly@uow.edu.au

Abstract

Background: Food is one of the most frequently promoted commodities, and promoted foods are overwhelmingly unhealthy. Marketing normalizes unhealthy foods, creates a positive brand image, and encourages overconsumption. Limited research is available to describe the extent of food marketing to children on web-based media, and measuring actual exposure is challenging.

Objective: This study aims to monitor the extent of children's exposure to web-based media food marketing as an essential step in increasing the accountability of industry and governments to protect children.

Methods: Children aged 13-17 years were recruited from October 2018 to March 2019. Children recorded their mobile device screen for 2 weekdays and 1 weekend day any time they visited relevant web-based platforms. After each day, the participants uploaded the video files to a secure server. Promoted products were defined using the World Health Organization European Region nutrient profile model.

Results: The sample of 95 children uploaded 267.8 hours of video data. Children saw a median of 17.4 food promotions each hour on the internet. Considering the usual time spent on the internet on mobile devices, children would be exposed to a median of 168.4 food promotions on the web on mobile devices per week, 99.5 of which would not be permitted to be marketed based on nutrient profiling criteria. Most promotions (2613/4446, 58.77%) were peer endorsed and derived from third-party sources.

Conclusions: Exposure to brand content that is seemingly endorsed by peers or web-based communities likely heightens the effects of marketing on children. Regulations to protect children from this marketing must extend beyond paid advertising to paid content in posts generated through web-based communities and influencers.

(J Med Internet Res 2021;23(7):e28144) doi:10.2196/28144

KEYWORDS

food; beverage; marketing; online; digital

Introduction

Background

Protecting children from the impacts of unhealthy food and beverage marketing has been repeatedly identified at the highest levels of global policy agenda setting as a priority intervention for childhood obesity prevention. The report from the World Health Organization (WHO) from the Commission on Ending Childhood Obesity explicitly urged governments to regulate the marketing of unhealthy food to protect children from obesity and related noncommunicable diseases [1]. The commission called for the implementation of previous WHO recommendations to restrict the exposure and power (persuasive content) of food marketing across all platforms and settings where children gather [2]. Robust empirical and review evidence

XSL•FO
RenderX

indicates that children's exposure to food marketing leads to a cascade of effects, including food brand awareness, positive brand attitudes, and purchase and consumption behaviors [3]. As promoted foods and beverages are almost exclusively for products high in added fat, sugar, and salt (*unhealthy* foods) [4], this may contribute to excess energy intake and poor dietary patterns [5].

For more than a decade, evidence on children's exposures to, and the impacts of, food marketing has identified the increasing prominence of digital or web-based media marketing [6].

Almost all Australian adolescents aged 13-18 years (94%) have their own mobile screen devices, and three-fourths have a social media account [7]. Most primary school–aged children also have their own mobile devices (67%), whereas 1 in 6 has a social media account. Adolescents spend an average of 43.6 hours at home in front of screens each week, most frequently engaged in web-based social media [7]. This includes screen time for entertainment, communication, and education. Screen time is lower for younger children, with an average of 31.5 hours per week [7] but still greatly in excess of screen time recommendations. Australian guidelines advise limiting recreational screen time to no more than 2 hours per day for children and adolescents [8].

With the advancement of the internet as a social and participatory space, marketers have been able to target and engage users with personal communications, infiltrate web-based communities with brand content, and enable peer endorsement of brand messages [9]. This means that brands can be embedded in web-based content and distort boundaries between marketing and editorial, entertainment, and personal material, diminishing children's capacity to recognize marketing as a paid promotion [10-12]. Crucially, web-based marketing is tailored to the unique characteristics and preferences of users, using data analytics that include users' personal information, browsing history, geolocations, and social media engagement [13]. This *behavioral targeting* of marketing on the internet has major implications for various aspects, including children's privacy, the impact of marketing on children, web-based marketing regulation, and monitoring of children's web-based marketing exposure to describe and understand the problem [13-15]. The immersive and interactive nature of web-based media and related marketing likely means that the effects of web-based food marketing on children are greater than those of other offline marketing [9,16].

Research evidence on the impact of web-based food marketing on young people has predominately focused on digital games, specifically *advergames* or branded games [17]. A meta-analysis of experimental studies examining the effect of exposure to web-based branded games on children's attitudes, choices, or intake of unhealthy foods identified a significant positive effect of small to moderate size [18]. Less evidence is available on the effects of other forms of web-based food marketing. In an earlier cross-sectional survey by the authors, we identified that children who reported higher web-based engagement with food brands and content, particularly through web-based videos, were more likely to consume unhealthy foods and drinks [19]. A systematic review of the evidence on the effects of web-based marketing of risk-associated products, including unhealthy foods, alcohol, and tobacco, indicated an association with children and young adults' attitudes toward, and intended and current use of, these products [20]. In interactive social contexts, including social media, child-brand interactions may be more influential in influencing children to form preferences for, and perceptions of, brands [21].

Monitoring children's exposure to food marketing is necessary for engaging policy makers and civil society on the issue, holding industries accountable for their marketing practices, and measuring the effectiveness of any regulations and compliance [22]. The International Network for Food and Obesity/noncommunicable diseases Research, Monitoring and Action Support (INFORMAS) is an international consortium of public interest organizations and researchers that seeks to support the generation of monitoring intelligence to describe food environments, including food marketing, and related policy responses. Global monitoring of children's exposure to food marketing across media platforms and settings, including television, has been undertaken [23]. However, the individualized nature of behavioral targeting of web-based marketing complicates efforts to monitor children's exposure to food marketing on web-based media.

Objective

This study aims to quantify and describe children's exposure to food and beverage marketing during their time spent on the internet, including the types of foods and beverages promoted and the platforms from where exposures were derived. We also sought to describe the nature of promotions, including the extent to which these were found in paid advertising space, on food companies' own sites and pages, or transmitted through web-based social networks. The approach used to capture marketing exposures also allowed us to identify the extent of children's engagement or interaction with food promotions. We hypothesized that children would be exposed to a high volume of unhealthy food and beverage marketing in their usual web-based interactions, which exceeded the number of promotions that they see for healthy choices, and that a large proportion of marketing would be peer endorsed and skewed toward third-party sources, such as shared content and blogs. Children aged 13-17 years were selected for the study, as they were deemed to have sufficient cognitive capacities to undertake the web-based survey and monitoring aspects of this project and to comprehend the ethical and privacy considerations of participating. Adolescents are also key social media users and targets for web-based food marketing [13]. Adolescents are susceptible to unhealthy food marketing despite their increasing cognitive ability, and they may be more impulsive in their purchase decisions [24].

Methods

Sampling and Recruitment

The study was approved by the University of Wollongong Human Research Ethics Committee (HREC 2018/158). Children were recruited through the national adolescent survey panel of the market research agency McNair yellowSquares. This panel comprises parents of young people across Australia who have agreed to be contacted to participate in research studies

XSL•FO
RenderX

(approximately 15,000 panel members). Panel members with children aged 13-17 years were invited to indicate their interest in participating in this study. Interested parents and children were sent the participant information sheets and consent forms to both sign and return. Participants were then asked to complete a short prestudy questionnaire that assessed their eligibility to participate, along with collecting information on their usual time spent on the internet on a mobile device and also on desktop and laptop computers, split by weekdays and weekend days. To be included in the study, children needed to have at least one social media account, log on to social media at least once per day, and have access to a mobile device (phone or tablet) that was compatible with the screen recording apps or settings. Only one child per family was chosen for participation. Participants were recruited in 2 rounds to avoid the school holiday period—October to November 2018 and February to March 2019. A sample size of approximately 150 children was sought from a national population estimate of approximately 1.4 million adolescents [25]. This was based on a margin of error of 5% for estimates of the average number of exposures to unhealthy food or beverage web-based promotions per day with a conservative population variance of 1000.

Procedure

Piloting

The study required children to record and upload data on their internet use on mobile devices and complete pre- and poststudy questionnaires. The main study was preceded by a pilot study of 26 children. The pilot led to major changes in the recruitment strategy (eg, increasing compensation for participant time), participant tracking and reminders, data coding, and improvements to the data upload server.

Screen Recording

Each participant was asked to video record their mobile device screen for 2 weekdays and 1 weekend day anytime they went onto relevant web-based platforms or apps. Relevant platforms include social media websites or apps, video sharing websites or apps, or browsing on the internet. They were asked not to record their screen when they were using any banking platform, using personal messaging (eg, SMS, Facebook Messenger, WhatsApp, or personal messaging on Snapchat or Instagram), making phone calls, or browsing through photos in their device's gallery. Participants nominated which days they would record within a 2-week period of entry into the study. They were sent 3 SMS text messages on nominated days as a reminder to record their screens.

Participants were provided with detailed written instructions and an instructional video to complete the study screen recording and upload tasks. The recording process varied across mobile device operating systems. For Android devices, participants were asked to download an app called the *Lollipop screen recorder*. The iOS participants had to move the screen recording setting of the control panel of their device. Participants could turn the recording on and off through this app or setting. The device showed a symbol at the top of the screen to indicate that the screen was being recorded. This recording function captures all user actions on their mobile device, such as scrolling, typing, and clicking.

Data Upload

Each participant was sent a unique log-in link to the McNair yellowSquares web-based database to upload the data. This was a bespoke platform for uploading files, completing questionnaires, tracking participants' study progress, and communicating any data issues. After each day of recording, the participants were instructed to upload their video files to the database. Participants were encouraged to edit videos using the video editing function on their device and to remove any footage they did not want the researchers to view. Given the size of the video files and the number of uploads being attempted simultaneously, upload to the database experienced issues with slow uploads and file corruption (inoperable files). Consequently, midway through data collection, new participants were instructed to submit their videos using WhatsApp. WhatsApp uses end-to-end encryption and does not store messages on its own servers.

The participants' video uploads were monitored daily during the data collection period. Data were deemed to be acceptable if the total duration of uploaded videos for the day was at least 30% of the reported usual time on the web on mobile devices (for weekdays and weekend days separately; from the prestudy questionnaire). When participants failed to reach this threshold of recording, they were contacted by email and phone, given further instruction, and asked to complete a replacement day. Participants received Aus $50 (US $38), paid into their research panel account, if they completed all 3 days of data recording. They received Aus $20 (US $15) if they only completed 1 or 2 days of recording. Participants were included in the final sample if they had at least one acceptable weekday and one weekend day.

Although there was minimal risk involved in participation, some of the main ethical concerns in the project were related to potential risks to privacy and confidentiality. Measures were taken to protect the privacy of the participants and to ensure data security.

Coding of Video Data

At the end of the study, all video data were transferred to CloudStor, a secure cloud storage server. Each video was watched at least twice by 1 person from a pool of 3 trained research assistants. In the first viewing of the video, all food and beverage promotions (including food and beverage products, retailers, and services) were identified and coded. The second viewing focused on recording the length of time spent on different platforms. Only branded food promotions were captured, including branded products and packages, brand logos, and brand characters. To be included, promotions needed to be shown onscreen for a minimum of 1 second and at least half of the brand name or logo needed to be visible.

The coding frame captured both the frequency and duration (seconds) of promotions onscreen, the nature of these promotions, and any participant engagement. Promotions were classified according to the platform (app or website) on which they occurred and the extent to which participants engaged with

XSL•FO
RenderX

the promotion by *liking*, *sharing*, *commenting*, or *clicking* on a link. Promotions were also classified as *paid*, *owned*, or *earned* media [26,27]. Paid media includes promotions generated by the food company, which pays to place these on third-party platforms. Examples include banner advertisements, paid search advertisements, and sponsored posts on social media. Owned media includes food brands' own websites, blogs, and social media pages. Earned media refers to promotions that do not directly come from the brand but are shared by third parties through reviews, reposts, blogs, referrals, and word-of-mouth.

Promoted products were defined using the WHO Regional Office for Europe nutrient profile model [28]. The WHO model designates products as *not permitted* or *permitted* to be advertised to children based on the thresholds for negative nutrients and energy content. Marketing for food companies, retailers, and restaurants that do not promote specific food products are not covered by the model. As there were a large number of company brand–only promotions for food retailers and restaurants that could not be classified using the WHO model, we also used the INFORMAS food classification system for monitoring food promotions (Table 1). This system classifies food into 3 broad categories—core or healthy, noncore or unhealthy, and miscellaneous—and 37 smaller food groups. Food and beverage retailers, restaurants, and delivery services were variously classified as noncore or miscellaneous, depending on whether they promoted a specific product and the nature of that product.

Interrater reliability was assessed with each research assistant independently coding the video data of the same 6 participants (15 days). The intraclass correlation coefficient was calculated for absolute agreement between the raters, giving an intraclass correlation coefficient of 0.97, indicating excellent reliability. Reliability results were discussed among the research team, and all issues were resolved before continuing. Reliability testing helped to refine the coding rules about the threshold of time, and the visibility of the brand, onscreen for the promotion to be counted.

XSL•FO
RenderX

Table 1. Frequency of food and beverage promotions in sample recordings (N=4446).

Food category	Frequency of promotions, n (%)
Core or healthy foods	108 (2.43)
Plain breads, rice, noodles, and crackers	20 (0.45)
Fruits and fruit products without added fats, sugars, or salt; ≥98% fruit juices	19 (0.43)
Milks and yogurts (≤3 g fat/100 g), cheese (≤15 g fat/100 g), and alternatives	18 (0.4)
Bottled water	15 (0.34)
Low sugar or high fiber breakfast cereals (<20 g sugar and >5 g dietary fiber/100 g)	15 (0.34)
Meat and alternatives, including unsalted nuts, seeds, and their pastes	10 (0.22)
Vegetables and vegetable products without added fats, sugars, or salt	4 (0.08)
Low fat or salt meals: frozen or packaged meals (≤6 g saturated fat and <900 mg sodium per serve), soups (<2 g fat/100 g, exclude dehydrated), sandwiches, and mixed salads	3 (0.07)
Healthy snacks: based on core foods (<600 kJ and <3 g saturated fat and <200 mg sodium per serve)	3 (0.07)
Oils high in mono- or polyunsaturated fats	1 (0.02)
Noncore or unhealthy foods	2579 (58.01)
Chocolate and confectionery	539 (12.12)
Fast food restaurant or delivery service: unhealthy options	503 (11.31)
Sugar-sweetened beverages	435 (9.78)
Alcohol	244 (5.48)
Sweet breads, cakes and biscuits, and high-fat savory biscuits and pastries	165 (3.71)
Savory snack foods with added salt or fat include chips, extruded snacks, flavored popcorn, and salted or coated nuts	155 (3.48)
Local restaurant or delivery service: unhealthy options	142 (3.19)
Supermarket or retailer: unhealthy options	85 (1.91)
Ice cream and iced confection	84 (1.88)
Other high-fat or salt products include spreads with added salt, animal fats, high-fat savory sauces (>10 g fat/100 g), and soups (>2 g fat/100 g, dehydrated)	75 (1.68)
High-sugar or low-fiber breakfast cereals (>20 g sugars or <5 g dietary fiber/100 g)	34 (0.76)
Full cream milk and yogurts (>3 g fat/100 g) and cheese (>15 g fat/100 g, high-salt cheeses) and alternatives	32 (0.72)
Flavored or fried instant rice and noodles	37 (0.83)
Sweet snack foods include sugar-coated dried fruits or nuts and nut- or seed-based bars	14 (0.31)
Fruit juice or drinks with <98% fruit	13 (0.31)
Meat and alternatives processed or preserved in salt	12 (0.27)
High-fat or salt meals: frozen or packaged meals (>6 g saturated fat or >900 mg sodium per serve)	10 (0.22)
Miscellaneous	1759 (39.56)
Fast food restaurant or delivery service: no specific product	931 (20.94)
Local restaurant or delivery service: no specific product	365 (8.21)
Supermarket or retailer: no specific product	207 (4.66)
Local restaurant or delivery service: only healthier options	111 (2.49)
Tea and coffee	51 (1.15)
Dietary supplements and sugar-free gum	26 (0.58)
Fast food restaurant or delivery service: only healthier options	25 (0.56)
Supermarket or retailer: only healthier options	22 (0.49)
Recipe additions: include soup cubes, seasonings, and other sauces	19 (0.43)
Food manufacturer: no specific product	2 (0.04)

XSL•FO
RenderX

Pre-Post Questionnaires

Participants were sent a unique link to a web-based questionnaire at the start and end of the study. This captured data on their usual time spent on the web on mobile devices and on all devices on weekdays and weekend days, social media use (on which platforms they had accounts, number of people per pages they followed on each account, and number of food brands they followed), number of food or beverage brand apps they had on their device, and number of emails or SMS messages they received each week from food or beverage companies.

Analyses

Statistical analyses were conducted using SPSS for Windows, version 25 (IBM Corporation). Data were analyzed descriptively, including the types of promotions (*owned*, *earned*, or *paid*) and nature of promoted foods. The rates of promotions (promotions per hour) were calculated based on the number of promotions on a sampled day divided by the total relevant video duration for that day. Relevant time spent on the internet included that which captured web-based use, excluding personal messaging and banking. The average hourly rate over the 3 days for each person was then calculated and weighted by day type (weekdays and weekend days). The reported usual time spent on the internet was used to extrapolate the rates of marketing during the recorded period to weekly exposures. On the basis of visual inspection of the data and the Shapiro-Wilk test of normality, the rates of web-based food marketing did not meet normality assumptions, and therefore, medians and IQRs were reported. The Kruskal-Wallis one-way analysis of variance with post hoc test and Bonferroni correction was used to compare the median rates of promotions across web-based platforms. Negative binomial regression was used to identify factors associated with higher weekly exposure to promotions. Independent variables included participant age, usual weekly time spent on mobile devices, number of accounts followed on social media, number of food brands followed on social media, and number of food apps on mobile devices. Negative binomial regression was used because the distribution of the outcome had greater variability than expected under a Poisson distribution. The sample mean of the dependent variable (210.0) was substantially smaller than its variance (43,813.1), and the dispersion parameter was 0.593 with a 95% CI that did not include zero, indicating overdispersion.

Results

Sample Description

The final sample of 95 children uploaded 272.8 hours of recordings, of which 267.8 hours were relevant (captured web-based use, excluding personal messaging and banking). The study completion rate was 14.8% (95/644). Across the 2 rounds of recruitment, 736 people were disqualified based on the prescreening questionnaire. Furthermore, 429 people declined to participate or did not start the task after qualifying, 95 dropped out during the study, and 25 were excluded as they did not reach the 30% video upload threshold of reported usual time on the web on mobile devices. Across the 280 days of recordings captured, 23% (22/95) reached a threshold of 75%-100% of the usual recorded time spent on the internet, 45% (43/95) captured 50%-74% of the usual time spent on the internet, and 32% (30/95) captured less than 50% of the usual time spent on the internet. Table 2 shows the characteristics of the study participants. Approximately half of the sample lived in suburbs classified as having a high socioeconomic status, based on the Australian Bureau of Statistics Socioeconomic Indices for Areas. Children reported usually spending an average of 12 hours on the web each week on mobile devices and almost 30 hours on the web each week across all devices. Almost all children held accounts on Instagram; in addition, most also had accounts on Facebook, Snapchat, and music streaming apps.

XSL•FO
RenderX

Table 2. Sample description (n=95).

Child characteristics	Statistics
Age (years), mean (SD)	16.2 (1.07)
Usual weekly web-based media use mobile devices (hours), mean (SD)	12.1 (9.71)
Usual weekly web-based media use all devices (hours), mean (SD)	28.9 (18.36)
Sex, n (%)	
Male	32 (34)
Female	63 (66)
Socioeconomic status (n=92), n (%)	
Low	15 (16)
Medium	26 (27)
High	51 (54)
Social media platform users, n (%)	
Instagram	87 (92)
Facebook	69 (73)
Snapchat	68 (72)
Music streaming apps	68 (72)
YouTube	43 (45)
Twitter	29 (31)
Pinterest	25 (26)
Twitch	10 (11)
Follow food brands on social media	
Frequency (n=91), n (%)	43 (45)
Number followed, mean (SD)	0.8 (0.47)
Food apps on phone	
Frequency (n=91), n (%)	71 (75)
Number of apps, mean (SD)	1.9 (1.80)
Emails or texts from food brands (n=91), n (%)	
None	27 (28)
1-5 per week	51 (54)
6-10 per week	11 (12)
11 or more per week	2 (2)

Types of Promotions

Across the sample recordings, there were 4446 food and beverage promotions. Of these 4446 promotions, 2613 (58.77%) were earned media impressions, 732 (16.46%) were on media *owned* by the brand (apps, websites, and pages), and 1101 (24.76%) were paid advertisements. Earned media impressions were mostly seen in content from other nonbrand organization or community sites or pages (eg, meme pages; 2221/2613, 84.99% of earned impressions). A smaller number of earned media impressions were from content shared by a friend (242/2613, 9.26%) or celebrity endorsements (150/2613, 5.74%).

Promoted Foods and Beverages

The INFORMAS food classification system was used to describe the nature of foods and beverages, as a large number (n=1840) could not be classified using the WHO European nutrient profiling food categories. The highest proportion of promoted foods and beverages was noncore (2579/4446, 58.01%; Table 1). An additional 20.94% (931/4446) of promoted foods and beverages were for fast food restaurants or delivery services that did not promote a specific food or beverage product. The most frequently promoted foods and beverages were fast food restaurants or delivery services (all advertisements combined: 1459/4446, 32.82%), local restaurants or delivery services (all advertisements combined: 618/4446, 13.9%), chocolate and confectionery (539/4446, 12.12%), sugar-sweetened beverages (435/4446, 9.78%), and supermarkets or retailers (all

XSL•FO
RenderX

advertisements combined: 371/4446, 8.34%). The most frequently promoted food and beverage brands were McDonald's (416/4446, 9.36% of all promotions), KFC (184/4446, 4.14%), Coca Cola (137/4446, 3.08%), Uber Eats (135/4446, 3.04%), Starbucks (123/4446, 2.77%), Boost Juice (113/4446, 2.54%), Woolworths (supermarket; 96/4446, 2.16%), Nutella (84/4446, 1.88%), Kit Kat (73/4446, 1.64%), and Coles (supermarket; 73/4446, 1.64%). Together, these top 10 promoted food brands contributed to almost one-third of all promotions.

Rate of Food and Beverage Promotions

Children were exposed to a median of 17.4 food or beverage promotions each hour on the internet for a total duration of 1.3 minutes per hour (IQR 1-2; Table 3). This included a median of almost 10 earned media impressions per hour. The median rate of promotions for foods that would not be permitted to be marketed based on the WHO nutrient profiling model was 50 times higher than the rate of promotions for foods permitted to be marketed. The food categories with the highest rates of promotions were fast food restaurants or delivery services (company promotions), fast food restaurants or delivery services (promoting unhealthy choices), sugar-sweetened beverages, and chocolates and confectioneries.

Table 3. Weighted median rates of web-based food and beverage promotions per hour and by weekly exposures on mobile devices.

Rate of promotions	Weighted median rate per hour (IQR)	Weighted median rate on mobile devices per week (IQR)
Total promotion count	17.4 (10-26)	168.4 (85-289)
Rate by media type		
Earned media	9.9 (6-15)	84.8 (40-177)
Paid media	3.7 (1-8)	36.1 (12-75)
Owned media	0.6 (0-3)	5.3 (0-36)
Rate by World Health Organization Nutrient profiling classification[a,b]		
Not permitted	10.0 (5-17)	99.5 (43-159)
Company brand only	4.4 (2-8)	37.2 (17-89)
Permitted	0.2 (0-1)	3.6 (0-8)
Not applicable	0.0 (0-0.4)	0.0 (0-5)
Rate by INFORMAS[c] food classification[d]		
Noncore foods	10.1 (5-17)	99.4 (43-159)
Miscellaneous	6.4 (2-10)	52.9 (24-99)
Core foods	0.0 (0-1)	0.0 (0-8)
Rate by food type[d]		
Fast food restaurants, no specific product	1.9 (0.6-4)	17.1 (6-46)
Fast food restaurants, unhealthy products	1.8 (0.3-4)	16.5 (5-34)
Chocolate and confectionery	1.5 (0.3-3)	12.4 (4-29)
Sugar-sweetened beverages	0.9 (0-3)	11.6 (0-27)

[a]Using World Health Organization for Europe Nutrient Profiling Model.

[b]2.59% (115/4446) could not be specified because of unavailable nutrition composition information.

[c]INFORMAS: International Network for Food and Obesity/noncommunicable diseases Research, Monitoring and Action Support.

[d]Using International Network for Food and Obesity/noncommunicable diseases Research, Monitoring and Action Support food classification.

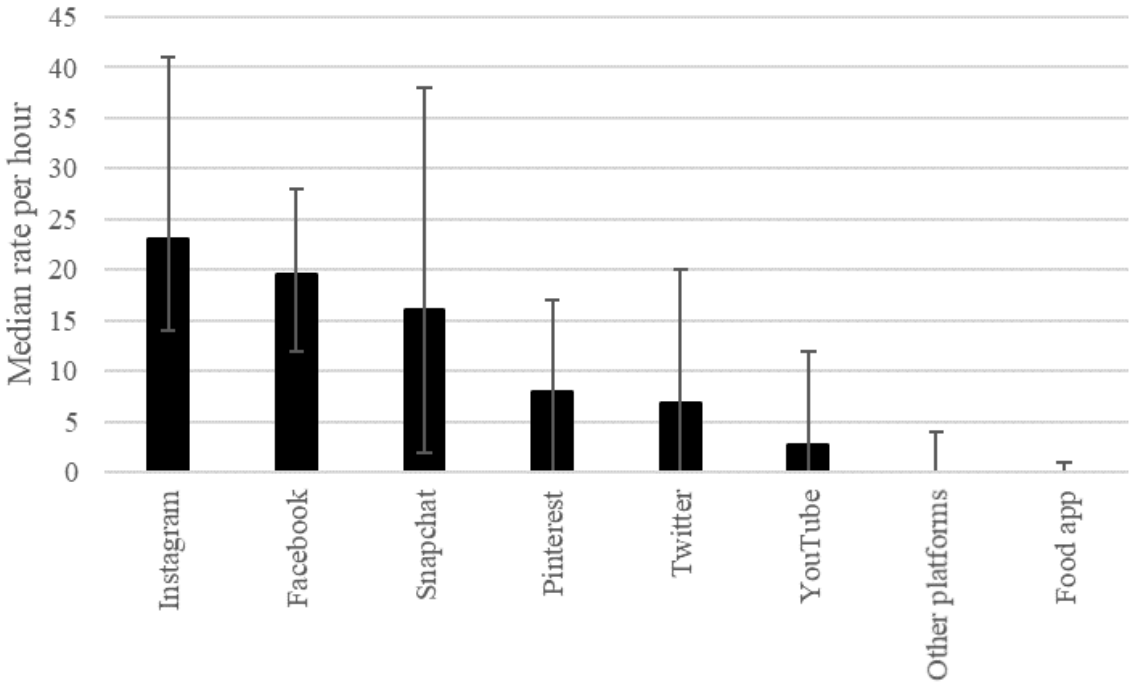
Considering children's reported usual time on the web on mobile devices, children would be exposed to a median of 168.4 food and beverage promotions on the internet on mobile devices per week for a total duration of 13.2 minutes (IQR 7-27). Children would be exposed to a median of 99.5 food promotions per week on their mobile devices that would not be permitted using WHO nutrient profiling criteria. This includes a median of almost 34 promotions per week for fast food restaurants or delivery services (company only or promoting unhealthy choices), 12.4 promotions for chocolate and confectionery, and 11.6 promotions for sugar-sweetened beverages.

The rates of promotions per hour varied by platform (Kruskal-Wallis H_7=142.12; P=.001; Figure 1). Post hoc pairwise comparisons with Bonferroni correction identified significantly higher rates of promotions on Instagram than on Pinterest, Twitter, YouTube, other platforms (apps and websites), and food apps (all values of P=.001). The rates of promotions on Facebook were significantly higher than those on YouTube, other platforms, and food apps (all values of P=.001). The rates of promotions on Snapchat were significantly

XSL•FO
RenderX

higher than those on other platforms and food apps (all values of *P*=.001).

Figure 1. Weighted median rates of web-based food and beverage promotions, by platform. The rates given as a function of time spent on platform, except for food apps, which is given as a function of total time on the web for only those reporting having food apps on phone. Error bars represent IQR. The number of participants visiting each site during the study was 76 for Instagram, 57 for Facebook, 40 for Snapchat, 11 for Pinterest, 22 for Twitter, 58 for YouTube, and 53 for other platforms (apps and websites).



Participant Engagement With Promotions

Participant engagement with promotions included *likes*, comments, shares, and clicking on links. Participant engagement with food and beverage promotions was low, with a median of 0.3 (IQR 0-2) overall engagements per hour of time spent on the web or one engagement approximately every 3.4 hours. Each week, participants were estimated to engage with brands a median of 3.9 times, although this was highly variable (IQR 0-21). In the highest decile of engagement, children were estimated to engage with brands at least 47 times per week.

Predictors of Exposure to Food and Beverage Promotions

Using negative binomial regression, the only factor that was significantly associated with weekly exposure to food and beverage promotions was the amount of time spent on the internet on mobile devices (B=0.54, SE 0.01; *P*<.001; 95% CI 0.03-0.08; Table 4). The incidence rate ratio (IRR) shows that for every 1 hour increase in usual time spent on the internet on mobile devices per week, exposure to food promotions increased by 6% (IRR=1.056, 95% CI 1.028-1.085; *P*<.001). Variables including age, number of accounts followed on social media, number of food or drink brands followed on social media, and number of food apps on mobile devices were not associated with exposure to promotions.

Table 4. Negative binomial regression incident rate ratios of the count of weekly exposures to food and beverage promotions on mobile devices.

Independent variable	Incidence rate ratio (95% CI)
Number of food apps	1.09 (0.98-1.21)
Usual weekly time on the web (mobile devices)	1.06 (1.03-1.08)[a]
Any accounts following on social media	1.02 (0.96-1.08)
Age	1.02 (0.85-1.24)
Food brands following on social media	0.95 (0.81-1.11)

[a]*P*=.001.

XSL•FO
RenderX

Discussion

Principal Findings

This study exposes Australian children's exceedingly high exposure to food marketing during their usual time on mobile devices. During each hour that a child spends on the internet on their mobile device, they would see more than 17 food and beverage promotions, equating to 168 promotions per week and 8736 promotions per year. For each hour increase in usual time on the internet on mobile devices per week, children's exposure to food promotions was found to increase by 6%. Our food marketing exposure estimates are likely to be highly conservative, given that they capture exposures only on mobile devices and not on desktop computers. There is some evidence to suggest that marketing on mobile and nonmobile devices is similar [29]; however, most social media use (where we observed the greatest rates of food promotions) is likely to occur on mobile devices [30]. In contrast, based on some of the most recent data on Australian children's exposure to food advertising on television, children were estimated to see around 4 food advertisements per hour during broadcast periods that attracted the greatest child audience, including 2 for unhealthy foods [31]. Considering that children aged 0-14 years watch commercial television for an average of 39 minutes per day [32], this would result in 3 exposures to food advertising on television each day or almost 19 exposures per week.

The rates of promotions for unhealthy products were far greater than promotions for healthier choices. Each week, children would be exposed to almost 100 promotions on their mobile devices for foods and beverages that would not be recommended to be marketed to children according to WHO nutrient profiling criteria. In addition, children would see around 17 promotions per week for fast food restaurant companies without a specific product promoted. These could not be appropriately classified using the WHO criteria as recommended to be permitted or otherwise, thus identifying a major limitation of these criteria for classifying food-related brands that should or should not be marketed to children. Although many fast-food outlets sell and promote *healthier* choices, observational data have identified that actual purchase of these healthier choices is infrequent [33]. Nutritional profiling of food companies that sell a range of products of varying nutritional quality, for the purposes of marketing policies, could be applied to the most frequently sold products.

We found that the greatest proportion of food promotion exposures earned media impressions. Although these promotions ostensibly derive from children's web-based social networks, the brand is often the initiator of earned media messages [34]. Almost all of the earned impressions were from either social media communities or celebrity pages, such as meme pages or web-based influencers. Posts from such third-party pages are increasingly sponsored by brands, allowing brands to access pages' huge networks of followers on the internet and social cache [35]. Experimental studies have identified that exposure to earned media impressions using web-based influencers leads children to consume more of a promoted snack compared with an alternative nonpromoted product [36]. The inclusion of advertisement disclosure in the web-based promotion had no effect on reducing the intake of the promoted snack. In Australia, the media industry introduced new standards in 2017, requiring marketing to be clearly distinguishable from other content, including promotions by web-based influencers [37]. Another review identified some evidence that exposure to, and engagement with, earned media impressions for alcohol positively increased intentions to consume alcohol and alcohol intake, whereas paid and owned media impressions had no effect [20].

Children in our study engaged with food and beverage promotions by *liking*, commenting or sharing content, or clicking on links in branded content. Although children engaged with relatively few of the overall branded impressions that they were exposed to, this would still equate to a median of almost 4 engagements with web-based food marketing on their mobile devices every week. It is worth noting that viewing impressions of web-based content is an important measure of reach, and even in the absence of engagement, is important to brands. Brands use a mix of promotions designed to increase reach, engagement, and click-throughs [38]. In an earlier cross-sectional survey of young Australian adults, participants' engagement with, but not their exposure to, web-based marketing for energy drinks was a predictor of their energy drink consumption [39]. The association between engagement with energy drink branded content and consumption in this earlier study was mediated by participants' subjective norms related to energy drinks, such as the perception that their peers frequently consumed energy drinks. The nature of earned media impressions on social media, which are shared through personal networks, likely boost such perceptions of product acceptability or use by peers [39]. Other qualitative studies have found that young adults are more distrustful of paid advertising on the internet but do not perceive earned media shared by friends as a form of marketing [11,12].

Our finding of the high rates of earned media impressions for unhealthy foods and beverages has major implications for public policy responses to protect children from this marketing. To inform new regulations planned in the United Kingdom to protect children from unhealthy food marketing on television and on the web, the government undertook an impact assessment to evaluate the potential costs and benefits of marketing restrictions [40]. This involved estimating current web-based food marketing exposures, which were based on marketing expenditure and principally considered banner advertisements. The number of advertising impressions (views by children) that were estimated using this approach was approximately five times greater for television advertising than for web-based marketing. Interrogations of the approach used by academics suggested that the impact assessment underestimated children's actual web-based food marketing exposures by 10-fold [41]. This study supports this finding, whereby Australian children's web-based food marketing exposures are approximately nine times higher than their exposure to food advertising on television. Our data also clearly refute the premise that banner advertising is a major source of web-based marketing.

To date, most studies seeking to assess the nature and extent of food marketing to children on the internet have been limited to

XSL•FO
RenderX

measuring either paid advertising on third-party websites [42] or owned media, including food company websites [43,44] or social media pages [45]. These studies are useful for capturing the range of techniques used by food companies and the techniques that generate the greatest overall user engagement. However, they are unable to quantify children's exposure to web-based food marketing or the extent of child engagement with brand promotions [9]. There are 2 notable earlier studies that assessed children's exposure to and engagement with food brand–related content on the web [27,46]. In one study, 101 Canadian children were recorded engaging in 2 social media apps for 5 minutes each on their own mobile device [46]. Similar to this study, based on the extrapolated study data, 12- to 16-year-old children were estimated to be exposed to food promotions an average of 189 times per week. The most promoted products were fast food and sugary drinks. Another study with 12- to 18-year-old from Belgium required participants (n=21) to capture screenshots of food images on their social media platforms over a 1-week period [27]. Aligned with our study, of the branded food images captured, approximately half were earned media impressions, including posts by web-based influencers and celebrities. Inspection of these earned media impressions in this earlier study suggested that these were likely paid marketing.

This study has some limitations. We did not achieve our target sample of 150 children. The final sample was lower than our original anticipated sample because of the substantial time involved in subject recruitment, technical errors with video uploads, and difficulties in obtaining complete data from participants. In an earlier pilot study, we had achieved a response rate of approximately 50%. This was substantially reduced in the main study, as we introduced a minimum threshold for daily video upload time. Surprisingly, we did not find significant associations between the number of overall accounts or food accounts that children followed on the internet or the number of food apps they had on their device and their marketing exposures. The CIs around the IRRs for these variables were wide, and future studies may be more adequately powered to detect significant associations. The minimum threshold for daily video upload of 30% of the usual time spent on the web on mobile devices meant that we did not capture all time spent on the internet on mobile devices. However, a comparison of the rates of food marketing exposures across data sets manipulated to include between 30% and 80% of the usual time spent on the internet recorded found there to be excellent reliability across the data sets (data not shown). Finally, the recruitment of children through the market research company survey panel may affect the generalizability of the findings to a broader population. However, this approach allowed us to capture a national sample, with representation from metropolitan and regional areas. This panel recruits multiple web-based and offline sources to recruit a broad spectrum of participants.

Opportunities for protecting children from web-based food marketing span legislative or regulatory controls, industry codes of practice for responsible marketing, and interventions that operate on an individual level to block exposure to marketing content. Internationally, some governments have introduced or are introducing restrictions on unhealthy food marketing to children on the web. As mentioned previously, the UK government announced plans to introduce a ban on all web-based marketing of unhealthy foods and beverages by 2022, as part of its national obesity prevention strategy [47]. In other jurisdictions, including Brazil, Peru, and Quebec, Canada, food marketing regulations apply comprehensively across all or most settings and media platforms, including digital media [48]. These latter regulations preclude advertising *directed to children*. The potential for marketers to circumvent this provision of restricting only marketing that is of specific appeal to children may limit the impact of such regulations.

To date, food industry codes of practice for responsible marketing largely fail to cover the types of web-based platforms that children use or the types of marketing they see or engage with on these platforms. For example, the International Food & Beverage Alliance Global Policy on Marketing Communications to Children only applies to media primarily directed to children aged <12 years and only applies to company-owned websites [49]. It does not, therefore, apply to any form of marketing on social media platforms, including company-owned pages. More recently, the digital media industry introduced responsible food marketing codes of practice. In October 2020, Google introduced restrictions on unhealthy food and beverage advertising for children aged <18 years in the United Kingdom and European Union [50]. This code requires advertisers to self-declare if they are using a web-based account to promote unhealthy foods or beverages, after which Google will block advertising from this account to children through its network of websites, videos, and apps. Although ostensibly this represents exceptional leadership by the media industry on this issue, previous restrictions on data mining and behavioral targeting of advertising to children aged <13 years on social media sites in the United Kingdom have been ineffective [51]. Given the global nature of Google, there is also no discernible reason why the policy should not be extended to all jurisdictions, rather than only where there is a threat or promise of government intervention. Our study identified significantly higher rates of exposure to food marketing on Instagram, Facebook, and Snapchat, signaling opportunities for these platforms to self-regulate to protect children from unhealthy food and beverage promotions. These platforms already self-regulate advertising content for tobacco products, although this is limited to restricting paid advertising. Research evidence shows that web-based promotion of tobacco products continues unabated through web-based influencers [52].

Finally, ad blockers and antitracking apps are available to block paid advertising and web-based tracking, which enables targeted advertising, on desktop computers and mobile devices. This includes software to block advertising and sponsored posts on social media. Some paid versions of social media platforms, such as YouTube Premium, also offer ad-free content. Although this study and others have highlighted that paid advertising is only a minority of the marketing impressions that children see on the web, this software may still be useful in reducing up to one-fifth of the web-based food marketing that children are exposed to. However, it is likely that widespread uptake of ad blockers would lead brands to invest further in earned and

XSL•FO
RenderX

owned media, thereby further increasing those types of media impressions.

Conclusions

Using real-time monitoring over a 3-day period, this study identified that Australian children are exposed to an outstanding volume of web-based food marketing on their mobile devices. This marketing is predominantly for unhealthy products and is shared through web-based communities. Children typically engage in web-based marketing multiple times each week. This exposure to, and interaction with, brand content that is seemingly endorsed by peers or web-based communities likely heightens the effects of marketing on children's brand attitudes and consumption behaviors. Governments and the media industry can and have designed policies to protect children from this marketing. The rapid acceleration and the use of data analytics and technologies used to capture personal data for targeted marketing is outstripping current legislation and policies for appropriate marketing regulations and related ethical concerns. To ensure that such policies are effective, they need to extend beyond paid advertising to paid content in posts generated through web-based communities, influencers, and celebrities.

Acknowledgments

This research was funded by the Australian Research Council Discovery Early Career Researcher Award (DE170100051). Piloting of the study was funded by an Australian National Preventive Health Agency grant (application ID1022310). The authors wish to thank Nina Balla, who contributed to the piloting of the data collection and analysis methods as part of her honors research project. The authors thank Amy Vassallo, who led the early testing of data collection methods and participant recruitment, and Grace Norton, who greatly contributed to the project through their management of data collection and data coding. The authors appreciate the time and effort given by the study participants in providing the video data.

Conflicts of Interest

None declared.

References

1. Report of the Commission on Ending Childhood Obesity. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/259349/WHO-NMH-PND-ECHO-17.1-eng.pdf [accessed 2021-06-15]
2. Set of Recommendations on the Marketing of Foods and Non-alcoholic Beverages to Children. World Health Organization. 2010. URL: https://apps.who.int/iris/bitstream/handle/10665/44416/9789241500210_eng.pdf?sequence=1 [accessed 2021-06-15]
3. Kelly B, King L, Chapman K, Boyland E, Bauman AE, Baur LA. A hierarchy of unhealthy food promotion effects: identifying methodological approaches and knowledge gaps. Am J Public Health 2015 Apr;105(4):e86-e95. [doi: 10.2105/AJPH.2014.302476] [Medline: 25713968]
4. Cairns G, Angus K, Hastings G, Caraher M. Systematic reviews of the evidence on the nature, extent and effects of food marketing to children. A retrospective summary. Appetite 2013 Mar;62:209-215. [doi: 10.1016/j.appet.2012.04.017] [Medline: 22561190]
5. Norman J, Kelly B, McMahon A, Boyland E, Baur LA, Chapman K, et al. Sustained impact of energy-dense TV and online food advertising on children's dietary intake: a within-subject, randomised, crossover, counter-balanced trial. Int J Behav Nutr Phys Act 2018 Apr 12;15(1):37 [FREE Full text] [doi: 10.1186/s12966-018-0672-6] [Medline: 29650023]
6. Cairns G, Angus K, Hastings G. The Extent, Nature and Effects of Food Promotion to Children: a Review of the Evidence to December 2008. World Health Organization. 2009. URL: https://apps.who.int/iris/handle/10665/44237 [accessed 2021-06-15]
7. Rhodes A. Screen Time and Kids: What's Happening in Our Homes? The Royal Children's Hospital Melbourne. 2017. URL: https://www.rchpoll.org.au/wp-content/uploads/2017/06/ACHP-Poll7_Detailed-Report-June21.pdf [accessed 2021-06-15]
8. Australian 24-hour Movement Guidelines for Children and Young People (5-17 Years). Australian Government Department of Health. 2019. URL: https://www.health.gov.au/health-topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians/for-children-and-young-people-5-to-17-years [accessed 2021-06-15]
9. Kelly B, Vandevijvere S, Freeman B, Jenkin G. New media but same old tricks: food marketing to children in the digital age. Curr Obes Rep 2015 Jan 10;4(1):37-45. [doi: 10.1007/s13679-014-0128-5] [Medline: 26627088]
10. Ali M, Blades M, Oates C, Blumberg F. Young children's ability to recognize advertisements in web page designs. Br J Dev Psychol 2009 Mar;27(Pt 1):71-83. [doi: 10.1348/026151008x388378] [Medline: 19972663]
11. Moraes C, Michaelidou N, Meneses RW. The use of Facebook to promote drinking among young consumers. J Mrk Manag 2014 Jun 3;30(13-14):1377-1401. [doi: 10.1080/0267257x.2014.909512]
12. Niland P, McCreanor T, Lyons AC, Griffin C. Alcohol marketing on social media: young adults engage with alcohol marketing on facebook. Addict Res Ther 2016 Nov 7;25(4):273-284. [doi: 10.1080/16066359.2016.1245293]

XSL•FO
RenderX

13. Tackling Food Marketing to Children in a Digital World: Trans-disciplinary Perspectives. World Health Organization Regional Office for Europe. 2016. URL: https://www.euro.who.int/__data/assets/pdf_file/0017/322226/Tackling-food-marketing-children-digital-world-trans-disciplinary-perspectives-en.pdf [accessed 2021-06-15]
14. Tatlow-Golden M, Garde A. Digital food marketing to children: exploitation, surveillance and rights violations. Global Food Security 2020 Dec;27:100423. [doi: 10.1016/j.gfs.2020.100423]
15. Montgomery K, Chester J, Nixon L, Levy L, Dorfman L. Big Data and the transformation of food and beverage marketing: undermining efforts to reduce obesity? Critical Public Health 2017 Oct 26;29(1):110-117. [doi: 10.1080/09581596.2017.1392483]
16. Boyland E, Thivel D, Mazur A, Ring-Dimitriou S, Frelut M, Weghuber D, On behalf of the European Childhood Obesity Group. Digital food marketing to young people: a substantial public health challenge. Ann Nutr Metab 2020;76(1):6-9 [FREE Full text] [doi: 10.1159/000506413] [Medline: 32101856]
17. Smith R, Kelly B, Yeatman H, Boyland E. Food marketing influences children's attitudes, preferences and consumption: a systematic critical review. Nutrients 2019 Apr 18;11(4):875 [FREE Full text] [doi: 10.3390/nu11040875] [Medline: 31003489]
18. Folkvord F, van 't Riet J. The persuasive effect of advergames promoting unhealthy foods among children: a meta-analysis. Appetite 2018 Oct 1;129:245-251. [doi: 10.1016/j.appet.2018.07.020] [Medline: 30031786]
19. Baldwin HJ, Freeman B, Kelly B. Like and share: associations between social media engagement and dietary choices in children. Public Health Nutr 2018 Dec;21(17):3210-3215. [doi: 10.1017/S1368980018001866] [Medline: 30086811]
20. Buchanan L, Kelly B, Yeatman H, Kariippanon K. The effects of digital marketing of unhealthy commodities on young people: a systematic review. Nutrients 2018 Jan 29;10(2):148 [FREE Full text] [doi: 10.3390/nu10020148] [Medline: 29382140]
21. Jones K, Glynn M. How children use social media for brand interactions. Young Consum 2019 Jun 17;20(2):91. [doi: 10.1108/yc-10-2018-0860]
22. Regional Action Framework on Protecting Children From the Harmful Impact of Food Marketing in the Western Pacific. World Health Organization Regional Office for the Western Pacific. 2020. URL: http://iris.wpro.who.int/handle/10665.1/14501 [accessed 2021-06-15]
23. Kelly B, Vandevijvere S, Ng S, Adams J, Allemandi L, Bahena-Espina L, et al. Global benchmarking of children's exposure to television advertising of unhealthy foods and beverages across 22 countries. Obes Rev 2019 Nov;20(Suppl 2):116-128 [FREE Full text] [doi: 10.1111/obr.12840] [Medline: 30977265]
24. Pechmann C, Levine L, Loughlin S, Leslie F. Impulsive and self-conscious: adolescents' vulnerability to advertising and promotion. J Pub Pol Mark 2018 Oct 18;24(2):202-221. [doi: 10.1509/jppm.2005.24.2.202]
25. Population by Age and Sex, Australia, States and Territories. Australian Bureau of Statistics. 2017. URL: https://www.abs.gov.au/ausstats/abs@.nsf/0/1CD2B1952AFC5E7ACA257298000F2E76?OpenDocument [accessed 2021-06-15]
26. Doyle C. Paid, owned, and earned media models. In: A Dictionary of Marketing. Oxford, UK: Oxford University Press; 2016.
27. Qutteina Y, Hallez L, Mennes N, de Backer C, Smits T. What do adolescents see on social media? A diary study of food marketing images on social media. Front Psychol 2019;10:2637 [FREE Full text] [doi: 10.3389/fpsyg.2019.02637] [Medline: 31824391]
28. WHO Regional Office for Europe Nutrient Profile Model. World Health Organization Regional Office for Europe. 2015. URL: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/publications/2015/who-regional-office-for-europe-nutrient-profile-model-2015 [accessed 2021-06-15]
29. Strout A. Mobile Vs Desktop: Are You Optimizing Both Experiences? Martech. 2016. URL: https://martech.org/mobile-vs-desktop-optimizing-experiences/ [accessed 2021-06-15]
30. Droesch B. More Than Half of US Social Network Users Will Be Mobile-only in 2019. Insider Intelligence eMarketer. 2019. URL: https://www.emarketer.com/content/more-than-half-of-social-network-users-will-be-mobile-only-in-2019 [accessed 2021-06-15]
31. Smithers LG, Haag DG, Agnew B, Lynch J, Sorell M. Food advertising on Australian television: frequency, duration and monthly pattern of advertising from a commercial network (four channels) for the entire 2016. J Paediatr Child Health 2018 Sep;54(9):962-967. [doi: 10.1111/jpc.13929] [Medline: 29660198]
32. Children's Television Viewing and Multi-screen Behaviour. Australian Communications and Media Authority. 2017. URL: https://www.acma.gov.au/sites/default/files/2019-08/Childrens-television-viewing-and-multi-screen%20behaviour.pdf [accessed 2021-06-15]
33. Wellard L, Glasson C, Chapman K. Sales of healthy choices at fast food restaurants in Australia. Health Promot J Austr 2012 Apr;23(1):37-41. [doi: 10.1071/he12037] [Medline: 22730936]
34. Edelman D, Salsber B. Beyond Paid Media: Marketing's New Vocabulary. McKinsey Quarterly. 2010. URL: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/beyond-paid-media-marketings-new-vocabulary [accessed 2021-06-15]
35. Brands Are Taking the Meme Route to Attract Teenagers. Los Angeles Times. 2019. URL: https://www.latimes.com/business/story/2019-11-27/brands-target-gen-z-with-memes [accessed 2021-06-15]

XSL•FO
RenderX

36. Coates AE, Hardman CA, Halford JC, Christiansen P, Boyland EJ. The effect of influencer marketing of food and a 'protective' advertising disclosure on children's food intake. Pediatr Obes 2019 Oct;14(10):e12540. [doi: 10.1111/ijpo.12540] [Medline: 31168959]
37. Clearly Distinguishable Advertising Best Practice Guideline. Australian Association of National Advertisers. 2016. URL: http://aana.com.au/content/uploads/2017/01/AANA_Distinguishable-Advertising-Best-Practice-Guideline__Final.pdf [accessed 2021-06-15]
38. Lee D, Hosanagar K, Nair HS. Advertising content and consumer engagement on social media: evidence from Facebook. Management Science 2018 Nov;64(11):5105-5131. [doi: 10.1287/mnsc.2017.2902]
39. Buchanan L, Yeatman H, Kelly B, Kariippanon K. Digital promotion of energy drinks to young adults is more strongly linked to consumption than other media. J Nutr Educ Behav 2018 Oct;50(9):888-895. [doi: 10.1016/j.jneb.2018.05.022] [Medline: 30297016]
40. Introducing a 2100-0530 Watershed on TV Advertising of HFSS (Food and Drink That Are High in Fat, Salt and Sugar) Products and Similar Protection for Children Viewing Adverts Online. Department for Digital, Culture, Media & Sport. 2019. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786554/advertising-consultation-impact-assessment.pdf [accessed 2021-06-15]
41. Tatlow-Golden M, Parker D. The devil is in the detail: challenging the UK department of health's 2019 impact assessment of the extent of online marketing of unhealthy foods to children. Int J Environ Res Public Health 2020 Oct 02;17(19):7231 [FREE Full text] [doi: 10.3390/ijerph17197231] [Medline: 33023267]
42. Vandevijvere S, Sagar K, Kelly B, Swinburn B. Unhealthy food marketing to New Zealand children and adolescents through the internet. N Z Med J 2017 Feb 17;130(1450):32-43. [Medline: 28207723]
43. Bragg MA, Eby M, Arshonsky J, Bragg A, Ogedegbe G. Comparison of online marketing techniques on food and beverage companies' websites in six countries. Global Health 2017 Oct 26;13(1):79 [FREE Full text] [doi: 10.1186/s12992-017-0303-z] [Medline: 29073900]
44. Neyens E, Smits T. Empty pledges: a content analysis of Belgian and Dutch child-targeting food websites. Int J Health Prom Educ 2016 Aug 19;55(1):42-52. [doi: 10.1080/14635240.2016.1218295]
45. Vassallo AJ, Kelly B, Zhang L, Wang Z, Young S, Freeman B. Junk food marketing on Instagram: content analysis. JMIR Public Health Surveill 2018 Jun 5;4(2):e54 [FREE Full text] [doi: 10.2196/publichealth.9594] [Medline: 29871854]
46. Potvin Kent M, Pauzé E, Roy E, de Billy N, Czoli C. Children and adolescents' exposure to food and beverage marketing in social media apps. Pediatr Obes 2019 Jun;14(6):e12508 [FREE Full text] [doi: 10.1111/ijpo.12508] [Medline: 30690924]
47. Tackling Obesity: Empowering Adults and Children to Live Healthier Lives. UK Department of Health and Social Care. 2020. URL: https://www.gov.uk/government/publications/tackling-obesity-government-strategy/tackling-obesity-empowering-adults-and-children-to-live-healthier-lives#shaping-the-marketing-to-our-children [accessed 2021-06-15]
48. NOURISHING policy database. World Cancer Research Fund International. 2020. URL: https://policydatabase.wcrf.org/level_one?page=nourishing-level-one#step2=3 [accessed 2021-06-15]
49. Nourishing Policy Database: Responsible Marketing to Children - Global Policy on Marketing Communications to Children. International Food and Beverage Alliance. 2014. URL: https://ifballiance.org/publications/responsible-marketing-commitments/?wpdmdl=2105&refresh=5fcec318e31341607385880 [accessed 2021-06-15]
50. Update to Other Restricted Businesses Policy. Google. 2020. URL: https://support.google.com/adspolicy/answer/9919030?hl=en-AU [accessed 2021-06-15]
51. Kids for Sale: Online Advertising and the Manipulation of Children. Global Action Plan. URL: https://www.globalactionplan.org.uk/files/kids_for_sale_v2.pdf [accessed 2021-06-15]
52. Where There's Smoke. Campaign for Tobacco-Free Kids. 2018. URL: https://www.takeapart.org/wheretheressmoke/ [accessed 2021-06-15]

Abbreviations

INFORMAS: International Network for Food and Obesity/noncommunicable diseases Research, Monitoring and Action Support
IRR: incidence rate ratio
WHO: World Health Organization

XSL•FO
RenderX

Edited by R Kukafka; submitted 22.02.21; peer-reviewed by K Raine, JJ Wu; comments to author 18.03.21; revised version received 30.03.21; accepted 24.05.21; published 12.07.21.

Please cite as:
Kelly B, Bosward R, Freeman B
Australian Children's Exposure to, and Engagement With, Web-Based Marketing of Food and Drink Brands: Cross-sectional Observational Study
J Med Internet Res 2021;23(7):e28144
URL: https://www.jmir.org/2021/7/e28144
doi:10.2196/28144
PMID:34255675

©Bridget Kelly, Rebecca Bosward, Becky Freeman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Blockchain for Increased Trust in Virtual Health Care: Proof-of-Concept Study

Anton Hasselgren[1], MSc; Jens-Andreas Hanssen Rensaa[2], MSc; Katina Kralevska[2], PhD; Danilo Gligoroski[2], PhD; Arild Faxvaag[1], MD, PhD

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:
Anton Hasselgren, MSc
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Mellomila 71
Trondheim
Norway
Phone: 47 46948498
Email: anton.hasselgren@ntnu.no

Abstract

Background: Health care systems are currently undergoing a digital transformation that has been primarily triggered by emerging technologies, such as artificial intelligence, the Internet of Things, 5G, blockchain, and the digital representation of patients using (mobile) sensor devices. One of the results of this transformation is the gradual virtualization of care. Irrespective of the care environment, trust between caregivers and patients is essential for achieving favorable health outcomes. Given the many breaches of information security and patient safety, today's health information system portfolios do not suffice as infrastructure for establishing and maintaining trust in virtual care environments.

Objective: This study aims to establish a theoretical foundation for a complex health care system intervention that aims to exploit a cryptographically secured infrastructure for establishing and maintaining trust in virtualized care environments and, based on this theoretical foundation, present a proof of concept that fulfills the necessary requirements.

Methods: This work applies the following framework for the design and evaluation of complex intervention research within health care: a review of the literature and expert consultation for technology forecasting. A proof of concept was developed by following the principles of design science and requirements engineering.

Results: This study determined and defined the crucial functional and nonfunctional requirements and principles for enhancing trust between caregivers and patients within a virtualized health care environment. The cornerstone of our architecture is an approach that uses blockchain technology. The proposed decentralized system offers an innovative governance structure for a novel trust model. The presented theoretical design principles are supported by a concrete implementation of an Ethereum-based platform called VerifyMed.

Conclusions: A service for enhancing trust in a virtualized health care environment that is built on a public blockchain has a high fit for purpose in Healthcare 4.0.

(J Med Internet Res 2021;23(7):e28496) doi:10.2196/28496

KEYWORDS

blockchain; ethereum; decentralization; Healthcare 4.0; virtualization; trust

Introduction

Overview

As a result of health care development, societies are undergoing a current demographic shift—people live longer, and fewer are born. The overall increase in life expectancy between 1970 and 2013 was 10.4 years on average for Organization for Economic Cooperation and Development countries [1]. A direct effect of this demographic shift [2,3] is that noncommunicable and chronic diseases become more prevalent, which presents a substantial socioeconomic challenge. Consequently, fewer

XSL•FO
RenderX

caregivers need to support an ever–increasing number of retirees with a rising number of chronic diseases. This unsustainable scenario is the strongest motivation behind many different ongoing proposals for transformations in the health care industry. Delivering health care, as we know it today, will most likely be unaffordable for any health system in 15 years from now, and many health services will have to be delivered by nonprofessionals and machines. This includes artificial intelligence health workers and devices connected via machine-to-machine protocols and automated, computerized services, which will be accessible via fast connections from anywhere, anyhow, and at any time (5G).

Furthermore, individuals will be forced to take more responsibility for their own health, take preventative measures, seek proper care in a timely manner, and behave more like autonomous patients. To facilitate this, there is a need to provide the right tools to encourage and enforce this transformation, both from the delivery side (health care providers) and the receiver side (patients). This transformation toward Healthcare 4.0 will challenge many of the present key components in a functional health system, where the concept of trust is one.

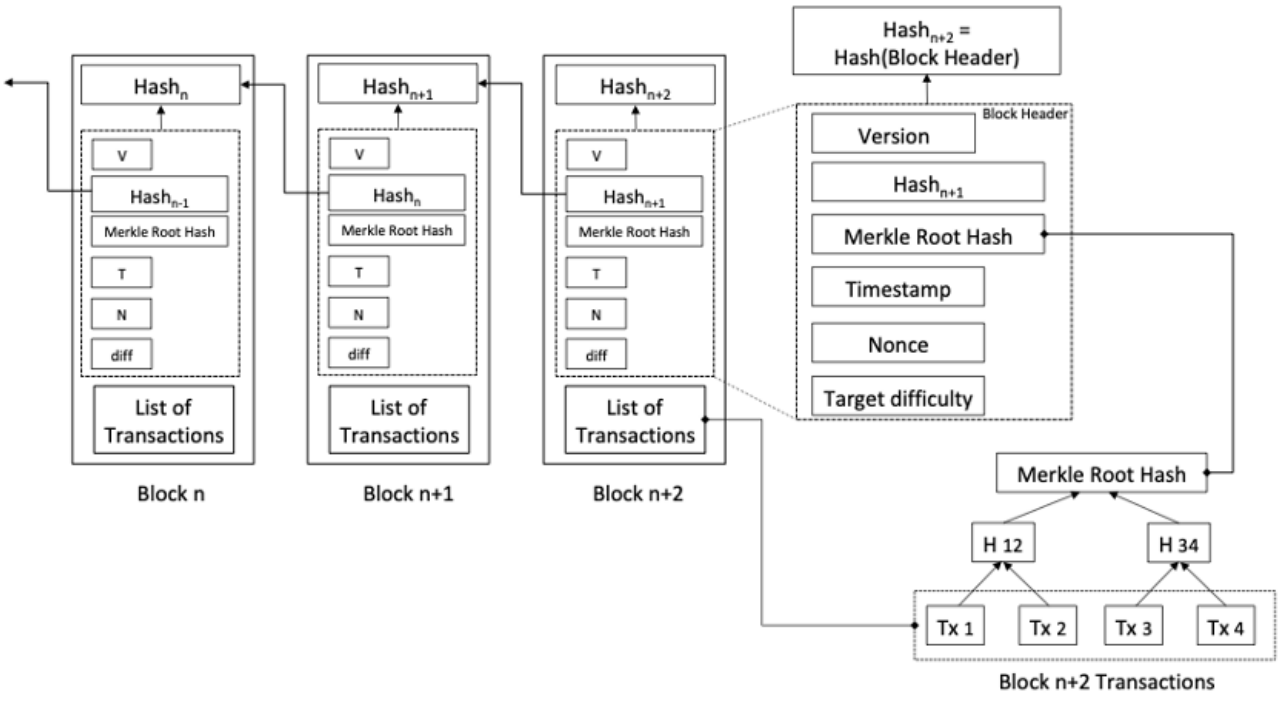
The first contribution of this paper is to review and predict the evolution of health care, and to identify the potential problems that could emerge in this transformation. It forms a theoretical foundation and urges the need for novel solutions to enhance trust. Second, the presented theoretical design principles are supported by the concrete implementation of a proof of concept. For this contribution, we choose the cornerstone in our architecture to be a blockchain technology implemented as an Ethereum-based platform called VerifyMed.

The remainder of this paper is organized as follows: the first section introduces blockchain and previous related work; the next section presents the method applied in this work; the following section outlines the results of technology forecasting and presents a trust issue in a virtualized health care environment; *the next* section presents a novel blockchain-based trust model for competence verification of health care personnel, and the final section provides a discussion and conclusions of the work.

Related Work and Blockchain Overview

Blockchain can be seen as an unconventional platform that alleviates the reliance on a single, centralized authority, yet it still supports secure and pseudoanonymous (or anonymous) transactions and agreements directly between interacting parties. It offers various degrees of decentralization, immutability, and consensus firmly founded in the mathematical principles of modern cryptography. A blockchain can also be described as an immutable ledger that logs data entries in a decentralized manner. In its original form, a blockchain enables entities to interact without a central trusted third party. The blockchain consists of a continuously growing set of data entries bundled together into blocks of data (Figure 1). Upon acceptance of the blockchain, these blocks are linked to the previous and future blocks sequentially [4]. In blockchain's original definition, this ledger of data blocks is decentralized and distributed across many nodes. This distributed ledger is transparent, verifiable by all, and tamper-proof. Owing to these properties, the blockchain has gained much attention for various applications. The first use case of a blockchain, Bitcoin, was introduced by a person or a group under the name of Satoshi Nakamoto in 2008 [5]. Bitcoin is also known as *a cryptocurrency.* Although cryptocurrencies remain the primary use case for blockchain, there is a substantial interest in applying this technology for other purposes and sectors [6]. Additionally, a blockchain allows for smart contracts—self-execution contracts that do not require any central authority. The use cases of blockchain in the health domain are increasing exponentially, as shown by Hasselgren et al [7], among others.

Figure 1. A generic overview of a blockchain structure.



XSL•FO
RenderX

Blockchain technology has five fundamental attributes that define the technology: (1) distribution, (2) decentralization, (3) time stamping, (4) data provenance, and (5) nonrepudiation. These five attributes are applied when addressing the fundamental problems in health care informatics and are part of driving the transformation toward Healthcare 4.0. The first generation of blockchain platforms led by Bitcoin [5] had a specially defined programming language for users to construct different transactions in the blockchain. The initial design rationale was that the programming language should be as simple as possible to satisfy the needs for building various transaction types and should not be a fully developed and powerful programming language. In computer science, the category of powerful programming languages is called the *Turing Complete*. The first blockchain platform that offered a Turing Complete language for programming, not just simple transactions but also more complex *smart contracts* and fully developed apps, was Ethereum [8]. There is an active debate on which concept is better and safer—development of malicious programs for blockchain platforms that do not have the Turing Complete programming language is very difficult and limited, in contrast to blockchain platforms that have the Turing Complete languages [5,9]. Nevertheless, it seems that the blockchain platforms that come with a fully developed Turing Complete programming language are very suitable for developing decentralized applications (dApps) for Healthcare 4.0, which is further elaborated in the next section.

Blockchain Platforms, dApps, and Smart Contracts

There are several decentralized platforms and frameworks for building dApps. Ethereum is the most common in health care applications [10]. This is most likely due to the large number of developers in the Ethereum community. Nevertheless, Ethereum has proven to be a solid platform for health care dApps [11]. Compared with the first and largest blockchain to date—Bitcoin—Ethereum has incorporated smart contracts, a function that substantially opens up the features of dApps built on Ethereum.

Smart contracts can be considered as self-executing contractual agreements, where preagreed upon provisions are formalized in the source code. Smart contracts can be automatically enforced based on these preagreed provisions, and they can work without any third party. The functions within a smart contract can be awoken in a blockchain transaction, and the use of this functionality could appeal to the health domain [8].

Zhang et al [11] stated that a well-designed health care dApp should limit the storage of encrypted sensitive data on the blockchain. Furthermore, they recommend that a dApp dealing with health care data should support Turing completeness to facilitate communication among various parties and handle the exchange of sensitive patient data. In the study by Kuo et al [10], there were clear indications that Ethereum, Hyperledger, and Multichain are more suitable platforms for the health domain than other blockchains.

Blockcerts [12] is a standard developed for verifying certificates of competence by storing signatures on a blockchain. The standard relies on existing trust relationships between the issuer and verifier of the certificate. Baldi et al [13] showed that certificates within this system could be spoofed. They also proposed the use of decentralized identifiers to govern such certificates. At present, there are private initiatives for medical credentials that use a blockchain. The first on the market was ProCredEx by Hashed Health [14]. They state that they have developed a blockchain-based solution that enables faster onboarding and credential verification.

Furthermore, a newly introduced collaborative project between Axuall, Inc and Metro-Health [15] announced a service for credentials of clinical practitioners. They state that they will enable digital portfolios that will include documentation of a practitioner's education, specialty training and board certifications, licenses, sanctions or medical malpractice judgments, evaluations, work history, and hospital affiliations. As these are private endorsers, there is no published peer-reviewed literature on their technical solutions. In addition to what is mentioned above, based on our knowledge, there is no published research that has addressed the same scope as our framework. As described earlier, several private organizations have created solutions for medical credentials by using blockchain technology. However, these are all based on the United States and are somewhat tailored to the US health system. We explore a broader solution in the form of a decentralized trust model that addresses the current issues with board certificates and credentials and creates an immutable portfolio of completed clinical work by a health care professional that is verifiable by all. The research approach used in this study is described in the following section.

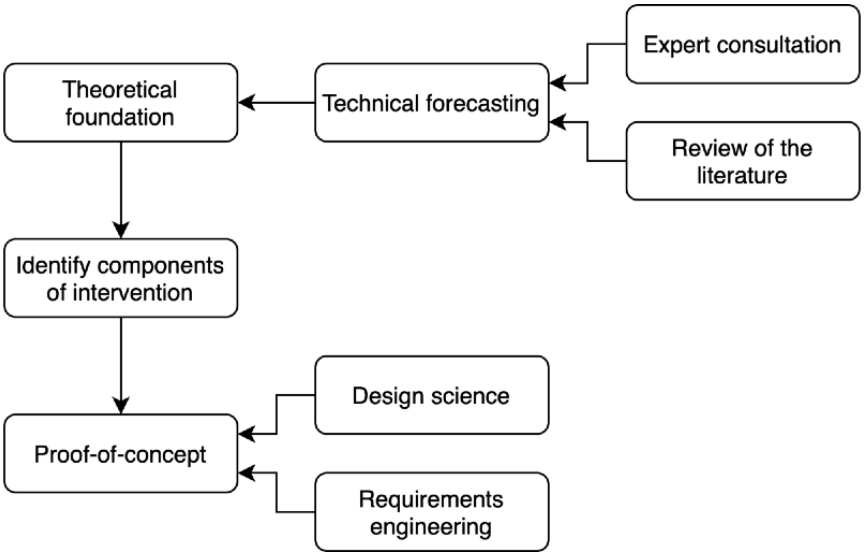
Methods

Overview

The research approach used in this study, as shown in Figure 2, follows one of the frameworks presented by Campbell et al [16], which describes a framework for the design and evaluation of complex intervention research within health care. Our study aims to address the following two key issues outlined in the framework: (1) establish the theoretical basis of the intervention and (2) identify and describe the components of the complex intervention.

XSL•FO
RenderX

Figure 2. Research approach for the presented work.



Summary of Knowledge and Technology Forecasting

We applied the most common method for addressing technology forecasting by reviewing the current literature and consulting domain experts [17]. The domain expert consultation was conducted in an unstructured manner; a convenient sample of (health) informatics experts from Norway was consulted about their views on the future of health care and Healthcare 4.0. A scoping review of the literature on the future of health care was performed in a semisystematic manner, and this is described in the section *Summary of Knowledge: Healthcare 4.0*.

Identifying Components of the Intervention

On the basis of the forecasting of Healthcare 4.0, a potential trust challenge is described and elaborated as the primary component of the intervention. This is presented in section *Trust in Healthcare 4.0*.

Proof of Concept

Furthermore, our work presents some technical components of a proof of concept to conceptualize (1) and (2), following the principles of design science [18] and requirements engineering [19]. This is presented in the section *VerifyMed: A Novel Trust Model*.

Results

Overview

We first describe a technology forecast of health care and then demonstrate how trust will emerge in this transformed health care system as a component of an intervention. The proof-of-concept VerifyMed is presented in a separate section, that is, *VerifyMed: A Novel Trust Model*.

Summary of Knowledge: Healthcare 4.0

Healthcare 4.0 [20] is a strategic concept for the health domain derived from the Industry 4.0 concept. The aim of Healthcare 4.0 is to allow for advanced virtualization to enable the personalization of health care in real time for patients, professionals, informal health workers, and nonhuman health workers. A transformation toward Healthcare 4.0 will be a shift from hospital or professional-centered health care (patient in hospitals) to a globalized, virtualized, and self-administered health care via distributed patient-centric care (multiple care providers) and later to patient-driven care fueled by personally generated health data.

Lasi et al [21] define Industry 4.0 with a wide range of current concepts: smart factories, cyber-physical systems, self-organization, new systems in distribution and procurement, new systems in the development of products and services, adaptation to human needs, and corporate social responsibility. Similarly, this categorization has been applied to health system development in the Healthcare 4.0 concept.

Thuemmler and Bai [20] state that:

> *The aim of Healthcare 4.0 is to allow for progressive virtualization in order to enable the personalization of healthcare next to real-time for patients, professionals, and formal and informal caregivers. The personalizing of healthcare will be achieved through the massive use of cyber-physical systems, cloud/edge computing, the Internet of everything including things, services and people and evolving mobile communication networks (5G).*

The six design principles from Industry 4.0 could be applied to Healthcare 4.0 to forecast health care transformation and to design applications with a high fit for purpose. The following design principles were proposed [22]:

1. Interoperability: enable people and machines to communicate through data standards and standardized infrastructure.
2. Virtualization: technologies for interoperability, faster internet connections, and connected devices enable the movement of parts of the physical processes in health care to a virtual environment.
3. Decentralization: linking real-time data and users together opens up more autonomous decisions and reduces the necessity of centralized services.
4. Real-time capability: a higher proportion of connected devices and people enables changes in real time.

XSL•FO
RenderX

5. Service orientation: a shift from products to services based on accumulating data could adapt faster to market changes.
6. Modularity: a higher degree of module-based delivery and configuration enables faster adoption of changing needs.

From an academic perspective, design principles are the foundation of the design theory [23]. As outlined in the section *Proof of Concept*, the design theory method is followed in developing our proof-of-concept platform, VerifyMed.

The following section presents an emerging problem in Healthcare 4.0, which serves as the basis for the components of our intervention.

Trust in Healthcare 4.0

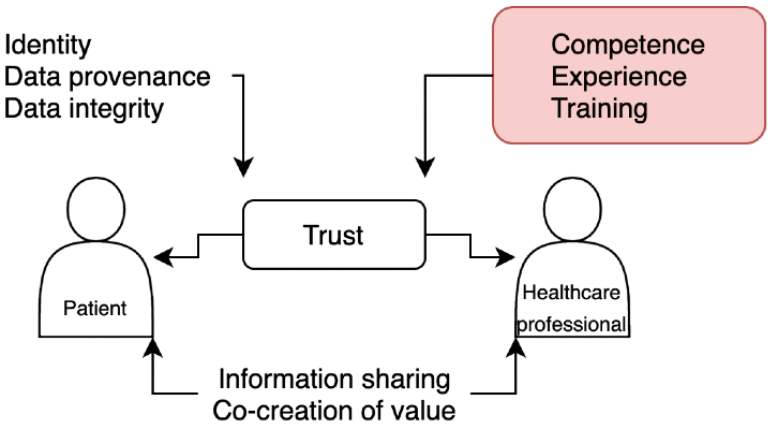
Overview

The definition of trust is a broad, multilayered, and complex concept that varies depending on the academic discipline that uses the term [24]. For this study, we have adopted the following broad definition of trust: *a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another* [25].

Trust From a Human Psychology Perspective

A central part of clinical practice is trust between a patient and a health care professional [26]. Maintaining trust with patients is a core function for physicians in their clinical practice [27].

Figure 3. Factors influencing trust in a patient-caregiver interaction.



Trust in Medical Technologies

When trusting medical technologies, institutional trust and technical reliability are deeply intertwined [37]. A key takeaway when reviewing Industry 4.0 is the need to explore further and understand how to build trust in the context of digital and virtualized health. This is related to trust in systems and information (human system) and people having the control of sharing information (human-human through the system).

Trust Issues in Healthcare 4.0

To conceptualize one part of the trust ecosystem in health care, we present the following theoretical issues with trust in a virtual patient-caregiver relationship: the patient needs to trust that the caregiver has the right competence (and authority) to deal with his or her health problem in a physical as well as in a virtualized health care environment. The caregiver needs to show the patient that he or she possesses the right competence to deal with the health problem of that specific patient; otherwise, the patient will possibly go somewhere else.

The General Medical Council states “patients must be able to trust doctors with their lives and health” [28]. This is also a part of the obligations of other health care professionals such as nurses [29]. Trust in health care professionals is considered a foundation for effective service delivery [30] and a core attribute in patient-centered care [31].

Commonly, trust is divided into interpersonal, social, and dispositional trust [32].

Furthermore, trust between a trustor and a trustee is encouraged by the trustee’s reliability (good reputation), competence (having skills to perform the task at hand), and integrity (honesty) [33]. Trust in a physician is related to increased treatment adherence, patient satisfaction, and improved health status [34]. Patients most commonly base their trust on doctor’s characteristics such as competence, compassion, privacy and confidentiality, reliability and dependability, and communication skills [35].

We know from other industries that a successful web-based consultation in health care delivery service requires a value cocreation between the caregiver and the patient [36]. Caregivers need active participation from patients to benefit from this cocreation. Several factors contribute to the trust foundation, which is the basis for value creation, as illustrated in Figure 3. Our approach targets the verification of competence, experience, and training (highlighted in Figure 3).

There are currently few or no systematic and objective tools to verify the competence and experience of health professionals in a transparent and accessible manner. The records of cases of delivered care are often stored in the electronic health record of the respective hospital. If a health care worker changes an employer, there is little or no opportunity to bring the ledger of given care (work experience). Like other industries, the health care industry has experienced a fast turnover of personnel. More health care workers change employers at a faster rate [38]. More health care workers are also moving across borders and jurisdictions at an increasingly higher pace [39]. In these cases, a tamper-proof, accessible record of the work history of someone as a health care professional, owned and controlled by no single entity, could be valuable. If this *portfolio* was stored in a decentralized manner, easily accessible with the consent of the

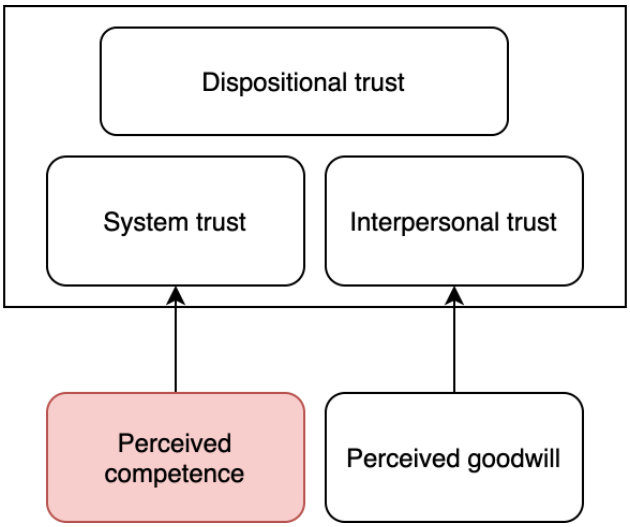
XSL•FO
RenderX

particular health care worker, onboarding processes for employers in the health care domain could be improved and that the health care worker could feel confident in that they control their own reputation by providing evidence-based care that could be verified at any time.

There is a need for patients, health care workers, and health care facilities to be able to verify the skill, competence, and formal certificates of health care personnel, especially when health care is moving toward Healthcare 4.0. Furthermore, it is essential to create an audit trail of complete work for health care workers; this could function as a portfolio that could potentially be used for future employers, freelance work, and increased confidence among health care workers.

Previous work has concluded that perceived competence and perceived goodwill are contributing factors to the system and interpersonal trust [24,32]. In a virtualized health care environment, it becomes increasingly important to verify the competence and credentials of health care professionals, as perceived competence is an essential component in building trust [40,41]. This highlighted component of perceived competence in Figure 4 is one part that the concept of VerifyMed partially addresses.

Figure 4. Trust model (adapted from Leimeister et al [32]).



The following section presents a proof of concept that addresses those needs.

VerifyMed: A Novel Trust Model

Overview

Our proposed architecture's technical core and the operational functionality are described in the studies by Rensaa et al [42-44]. In addition to that technical part, we describe some of the crucial functional and nonfunctional requirements and the principles that influenced our design rationale.

Our proposed architecture provides a solution for enhancing trust between a caregiver and a patient within a virtualized health care environment. The cornerstone feature in our architecture is its ability to capture trust relationships within the health care system and put them in a blockchain. Patients can use this trust mechanism to confirm credentials and potentially enhance trust in a caregiver during their interaction. Furthermore, the architecture includes tools for evaluating these interactions publicly on a blockchain. These evaluations served as a file for the caregiver's experience. We proposed the following three types of evidence for building trust in a virtualized health care environment: evidence of authority, evidence of experience, and evidence of competence [43].

The functional requirements describe the key features that we desire in our system based on our problem statement. Nonfunctional requirements describe the properties of the system, such as security, privacy, and performance requirements. Nonfunctional requirements often have a sizable architectural impact on how the system is implemented, whereas the functional requirements present the functionality that should be present within the architecture. These requirements are deduced from both industry requirements for handling patient data and the perceived problems deduced from our problem statement.

Previous research on blockchain apps within the health care industry has defined general principles for the requirements and system design principles that should be followed. Zhang et al [11] defined the metrics for evaluating blockchain apps within the health care industry. Although they are primarily directed toward the American Health Insurance Portability and Accountability Act (HIPAA), we generalize and try to capture some of these principles in our requirements.

Regulatory Compliance: Compliance With Current Health Data Laws and (Health) Privacy Regulations

Several regulatory bodies are responsible for preserving privacy and access rights to personal health data. The most prominent are the HIPAA for the United States and the General Data Protection Regulation (GDPR) for the European Union. In addition, most countries have national health data laws that further regulate health data for their citizens. In the scope of this study, we explored the GDPR compliance for VerifyMed. There are currently some uncertainties around general blockchain compliance with the GDPR [44], and these uncertainties, mainly around the level of anonymization and identification of data controllers in a decentralized network,

XSL•FO
RenderX

have not yet been clarified in any court case by the European Data Protection Board. However, it has been argued that there are no compliant blockchains, only complaint use cases, and apps [44]. The VerifyMed platform is designed to enhance user privacy and access control, and the following relevant GDPR articles have been addressed [45].

The VerifyMed platform is also designed to enhance the right of access by the data subject (Article 15 of the GDPR). As the system is designed not to store any personal data on the blockchain, it is also compliant with Article 17 of the GDPR (right to erasure or "right to be forgotten"), which only refers to personal data.

As the system is decentralized by design and there are possibilities for the user to access and receive the data at any time, it is compliant with Article 20 of the GDPR (right to data portability). The system requires an identity management solution to ensure full anonymization of the users and complies with Article 32 of the GDPR (security of processing). Identity management is not addressed within the scope of this study.

Key Functional Requirements

In accordance with the patient-centric health care system, we chose to define our main functional requirements in the context of the patient. As will be described later, the blockchain component of our architecture can be defined as a provider-centered model. We also note that fulfilling our patient-centered requirements allows the architecture to be used in settings outside of the patient and caregiver relationship. The main purpose of the model was to serve a patient-centered use case. The key patient-centered functional requirements were as follows:

1. Verification of caregiver credentials: a patient using a third-party system to talk with a caregiver should be able to verify the credentials by only using data from the blockchain. The patient must be able to do so without relying on any trust in the medical professional.
2. Verification of caregiver experience: a patient should be able to evaluate the experience of a medical professional by looking at data from the blockchain. Thus, the credibility of the data on blockchain must be enforced. The presented patient-centered requirements trigger opinionated system design choices to support this functionality. We additionally define two key features and refer to them as other deduced requirements. Therefore, these features will be subject to further specifications through nonfunctional requirements.
3. Transparency of blockchain data: to support data transparency to patients, we chose to use a publicly available blockchain to store the blockchain data. As these blockchains often have an associated fee with transactions, the system must take this into account.
4. Governance of blockchain data: to ensure that the trust relationships on the blockchain are anchored in the real world, they should be anchored in the existing corresponding trust relationships within the health care system. Just as there are governance entities responsible for credentials in the real world, they should be present in the proposed architecture as well.

Nonfunctional Requirements (via Quality Attributes)

Overview

In addition to the functional requirements above, we also surface the nonfunctional attributes of the system through quality attributes. The number of quality attributes of a system is unbounded. Therefore, this section presents the quality attributes that are considered to have the most significant architectural impact on the system.

Security Requirements

Fraudulent Treatments

A treatment cannot be published in the blockchain by unauthorized parties. All treatments must be cryptographically protected by an entity with direct or implicit authority to publish treatments.

Fraudulent Treatment Approvals

A treatment cannot be approved on the blockchain by unauthorized parties. All treatments must be approved by a license holder who the patient approves.

Fraudulent Evaluation

It should be impossible to publish an evaluation without going through a valid treatment. Once treatment has a related patient-reported outcome measure (PROM) published, it should not be possible to create another PROM related to the same treatment.

The Integrity of Treatments

It must be possible to ensure that a treatment or evaluation has not been tampered after their publication to ensure the credibility of these data sets. It is possible to prove this by using blockchain data.

Privacy Requirements

Unlinkability to Patients

The identity of patients must be treated as confidential. It should not be possible to link a transaction on the blockchain to a specific patient without any further knowledge from outside the blockchain. This will contribute to making the proposed system GDPR and HIPAA compliant (reference to regulatory compliance).

The Anonymity of Patients

The content of evaluations and treatments published on the blockchain should not reveal the identity of patients. The data published on the blockchain should either be a summary that cannot be linked to the patient or in another format that cannot be linked to a specific patient.

Access to Patient Data

The complete evaluations, including data linkable to patients, should be stored outside the blockchain. These data sets should be used to control patients. Access to these data sets for entities outside the patient and caregiver interaction should be denied unless the patient grants access.

XSL•FO
RenderX

Availability Requirements

Addition of New Governance Entities

It should be possible to add new governance entities dynamically without any code changes to the original contracts on the blockchain.

Recoverability After Authority Loss

If a governance entity becomes permanently unavailable or misbehaves, it should be possible to remove it, that is, to recover the dApp into a healthy state without interaction from the misbehaving authorizing entity.

Scalability Requirements

The amount of data on the blockchain should be minimal: the public blockchain is an expensive storage medium. Small data formats and encoding should be used to represent the data in the blockchain.

Performance Requirements

Minimization of transactions: interactivity with the blockchain should be reduced. The number of transactions required to go from the start to the published PROM should be small.

The Architecture

Overview

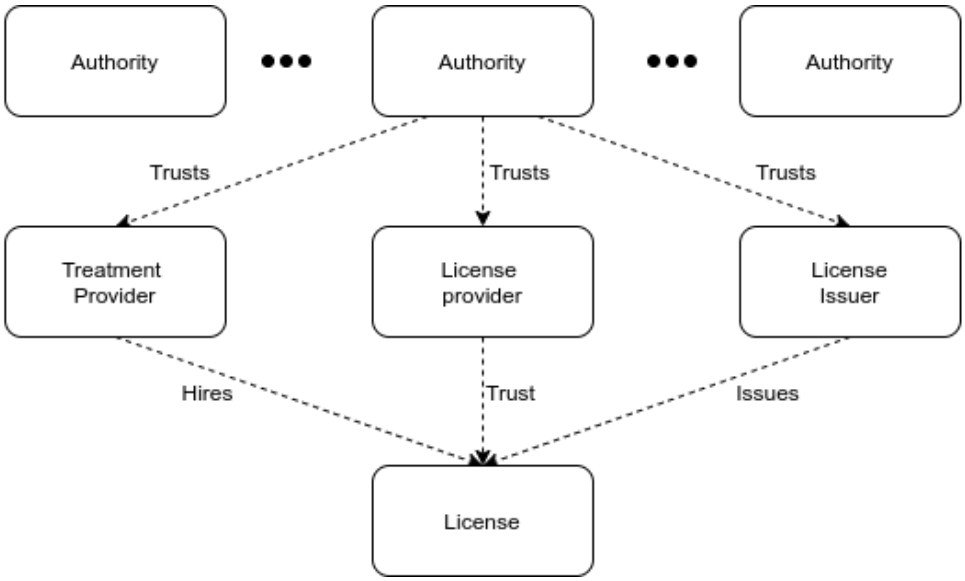
Our novel architecture provides trust between caregivers and patients within a virtualized health care environment. This is done through three main processes: evidence of authority, evidence of experience, and evidence of competence, each with its own components and stakeholders associated with them. We first define the terminology used in our architecture. Second, we present our proposition through an overall reference architecture. Finally, we describe how we further refine the reference architecture. We do this by describing the processes in the order in which they occur in the real world, along with the main components associated with them.

Terminology

Our architecture uses a concept for many different stakeholders, each represented by a given terminology. The stakeholders shown in Figure 5 are defined as follows:

1. Authorities: these are top-level government actors that have the overall responsibility of the health care sector (eg, national health directorates).
2. License: a license represents the practitioner's role as health personnel. Although a license in a traditional sense is the authorization of health personnel, we instead use it to represent the personnel themselves. Authorization is captured through trust relationships related to licenses.
3. License issuer: organizations that issue licenses for health personnel. License issuers are the only ones that can create licenses.
4. License provider: organizations that give formal authorization to practice for a license.
5. Treatment provider: organizations in which practitioners operate and are responsible for issuing treatments for patients. Examples include hospitals, clinics, and virtualized health care services.

Figure 5. Trust relationships published on the blockchain.



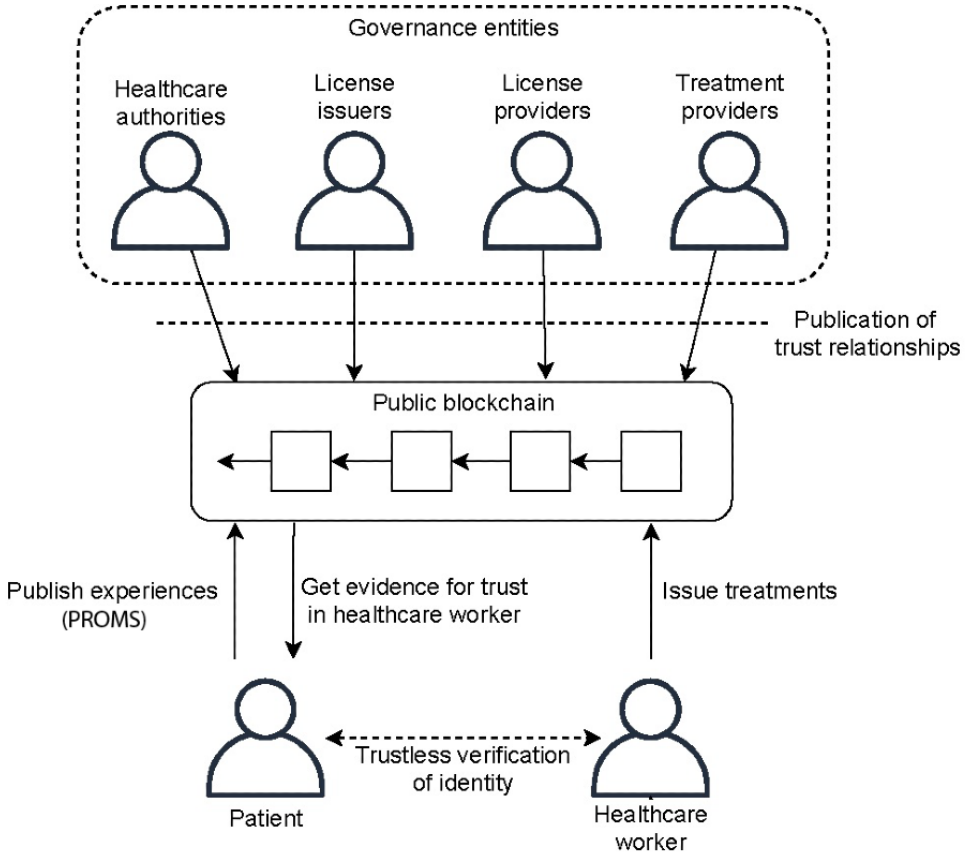
Overall Reference Architecture

Overview

As described in the functional requirements, the goal of VerifyMed is to provide trust in a health worker from a patient's perspective. The high-level reference architecture is shown in Figure 6. It captures trust by using a blockchain to store the formal trust relationship from health care organizations to health workers. Furthermore, as health workers issue treatments over time, summaries of these are published on blockchain. Finally, the evaluations of these treatments were published on the blockchain. The result is that the formal credentials of a practitioner can be validated through trust relationships, and their experience can be captured through logged treatments and evaluations.

XSL•FO
RenderX

Figure 6. The VerifyMed architecture for providing trust in a virtualized health care environment. PROMS: patient-reported outcome measures.



Creating Trust in a Caregiver

The first goal of the architecture is to capture the formal trust relationships between organizational actors and care providers within the health care industry. The end goal is to form a deduced trust relationship from health care authorities to the care provider and to capture the relationship in a way that is transparent and can be validated by the patients.

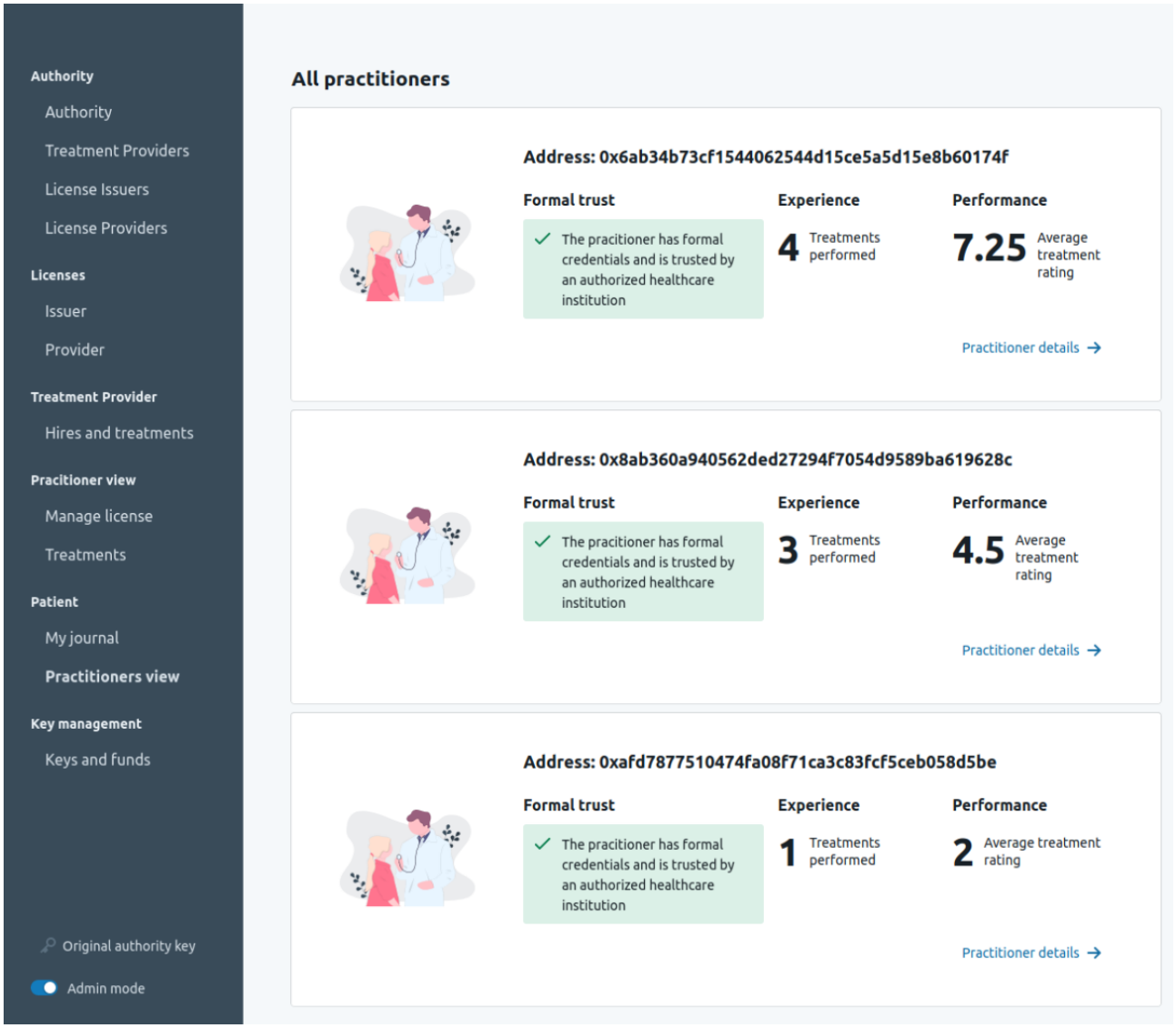
Figure 5 describes our model for trust relations between organizations and care providers. The top level was composed of large health care authorities. Authorities organize themselves through a model of distributed governance, for example, through simple voting, where existing trusted authorities can vote for the addition or removal of an authority. The main role of authorities is to provide trust in the defined stakeholders, who issue, authorize, and hire license holders. License holders can only practice and otherwise interact with the blockchain if all their upstream relations are linked to an authority. The patient entity is not part of this trust hierarchy; that is, patients are invited to publish PROMs on the blockchain by the care providers who have a trusted license after a completed treatment or interaction.

Caregiver and Patient Interaction

Once a license is considered trusted through the relationships captured on the blockchain, patients can use this information to check it. When meeting a practitioner, they can use the procedures defined in the smart contracts to check if their license is trusted and valid. Figure 7 illustrates the verification of the license, experiences, and skills of health practitioners.

XSL•FO
RenderX

Figure 7. Verification overview of health practitioners.



Evaluation of the Treatment

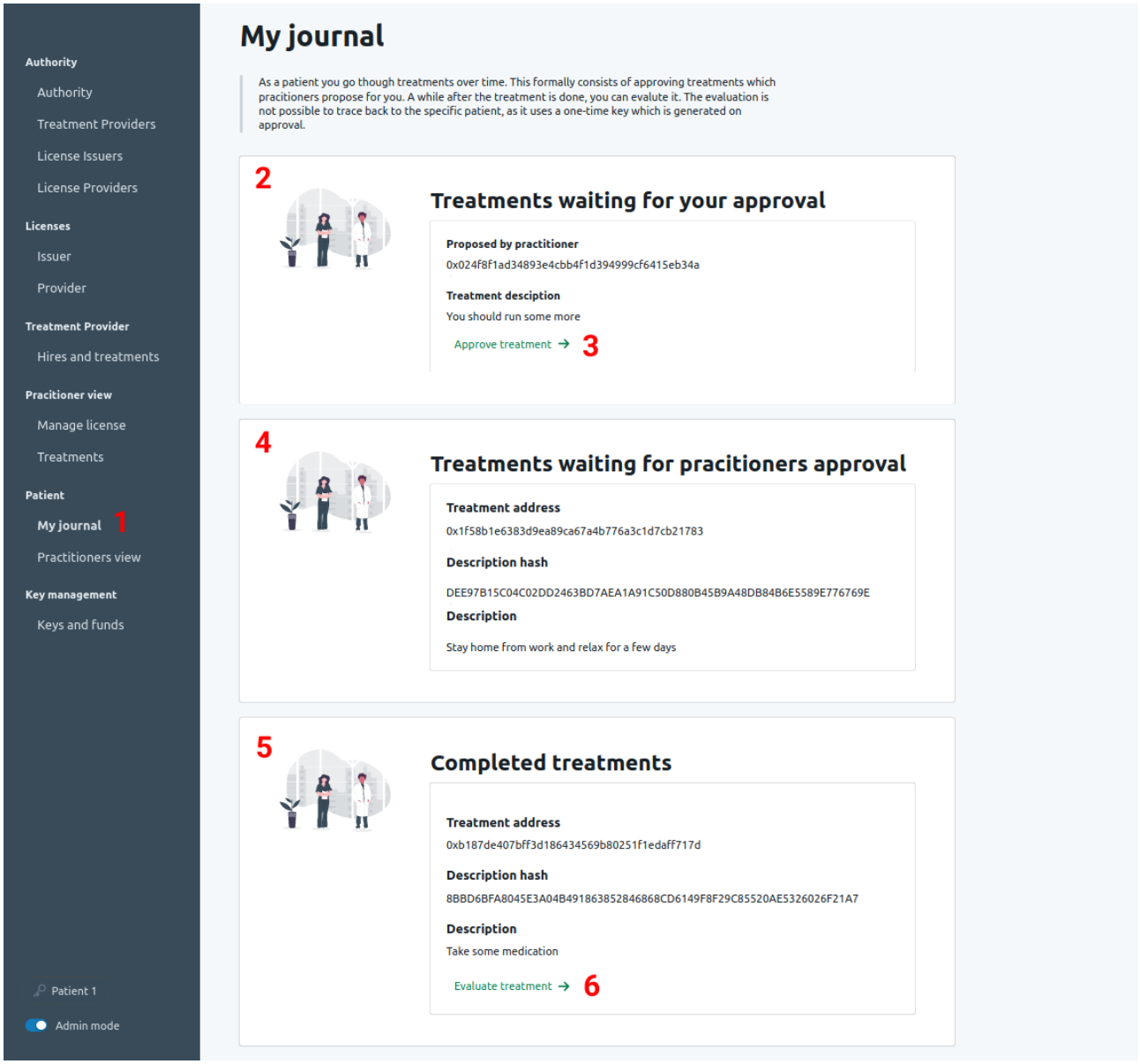
Once the treatment is completed, the patient may evaluate the treatment. The patient can do so via the one-time key generated during the treatment creation and thus create the evaluation without revealing their identity. This evaluation can be linked implicitly to a treatment provider and an approving practitioner. Future patients can use this information to enforce or decrease their trust in a practitioner.

Usage Outside of the Patient and Caregiver Relationship

Although we focus on the patient and caregiver relationship in the context of treatment, we also surface how public data sets have many use cases outside of this setting, such as audits, second opinion services, reporting, and evaluation of treatment providers. Figure 8 shows the user interface for the patients.

XSL•FO
RenderX

Figure 8. Overview of the user interface for patients.



Discussion

This work outlines a theoretical basis for the need for a blockchain enhanced trust model in a virtual health care setting, which contributes to the overall understanding of how the health care sector transforms into a new era, Healthcare 4.0, and the potential problems that could arise along with this transformation. Following this analysis, we built and implemented the novel VerifyMed platform that could trust in a virtualized health care environment.

We have used design principles from the Industry 4.0 concept to forecast Healthcare 4.0 and address an emerging problem in a future health care system. Our results show that our proof-of-concept implementation can be used to verify the authority of a health care worker, experience, and competence. The verifier does not have to place any trust in the health care workers themselves. This process can be performed by anyone with access to the Ethereum blockchain network, making the evaluation process fully transparent. In the further development of the system, microcredentialing can be incorporated, making it possible to verify specific skills among health care professionals.

Our trust model is justified in real-world governance of health care. As an environment with heavy regulatory oversight, capturing pre-existing governance relationships on a public blockchain serves as a natural first step for providing trust in virtualized settings. Furthermore, we strengthen our model by adding revocation abilities, where the trust of a governance entity can be revoked if it acts in bad faith. The result is a trust model justified in the inherited trust relationship between patients and the currently established health care system.

The VerifyMed platform enables individuals to store their respective credentials in a secure and accessible manner. The provenance of these data can be guaranteed using the immutability of the blockchain. In theory, this should mitigate the need to constantly verify the credentials from the issuing body and potentially speed up recruitment and onboarding processes in the health care sector.

XSL•FO
RenderX

We note that our trust model is extensible. A patient may trace all trust relationships from any evidence back to a top-level authority. The patient stands free to blindly trust the blockchain or use a third-party service to independently verify each of the upstream governance entities.

Health data are inherently sensitive, and thus, demand privacy. The management, storage, and access rights of health care data are highly regulated, both through general data protection acts such as the GDPR, health data specific acts such as the HIPAA, and often national health data laws. In an initial analysis [46], VerifyMed complies with the GDPR, although the general compliance of blockchain and the GDPR is under investigation; this work may have to be updated. Future work should include a comprehensive compliance analysis, and if appropriate, suggest an adaptation to comply with specific national health data laws and the HIPAA.

VerifyMed does not cover an identity solution for any of the users, and this is obviously an important component for the system to be ready for a real-life setting. As identity management is a core function in a health informatics system, future work must address this issue and develop an identity solution fitted to this particular use case.

VerifyMed uses the public Ethereum blockchain to host smart contracts. This choice is incorporated into the architecture, as the public nature of the blockchain is considered. Using a public blockchain requires limiting the published data to protect patient privacy, and access control schemes must be implemented within smart contracts. In addition, there is a need to incorporate a mechanism to transfer Ether (or smaller fractions of gas; ie, gwei or nanoeth) between accounts, thereby allowing them to submit transactions. The key advantage of using a public blockchain for this use case is transparency, no need for interorganizational agreements, and the possibility of interacting with the underlying cryptocurrency of Ethereum. The disadvantages of using the Ethereum blockchain are the monetary price of transactions and scalability issues related to low throughput. Furthermore, as the platform is governed by a set of authorities, license issuers, license providers, and treatment providers, this allows the publication of evidence for trust rooted in real-world trust relationships on the blockchain. This model contrasts with the fully trustless principles, which are usually applied within public blockchains but are necessary for the complex system of the health care domain. However, this can open up using a permissioned blockchain instead of fully public, which could have benefits such as reduced transaction costs and higher scalability. This should be explored in future studies.

VerifyMed could, with future updates, enrich the current trust model by including more trust requirements, such as (1) the caregiver must trust that the patient exists, (2) the caregiver must trust the authenticity of the data that the patient is willing to share, and (3) a third party (eg, an insurance company) must be able to trust the claim of the patient that care provision has taken place. The patient cannot really understand the credentials and experience of a caregiver because having a license is not the same as having credentials and having competency is not the same as having experience. Thus, the system should make the credentials or competency contextually important to the patient.

In the forecasting analysis, experts were consulted based on a convenience sample. This is not a comprehensive review of the general opinions of experts but just guidance in the direction of forecasting. It is not possible to preclude that this sample was not biased. However, a review of the literature supports input from expert consulting.

The trust mechanism that the blockchain enables in this concept provides a more transparent, accountable, and controlled handling of verifying competence and experience. This could also be achieved using a centralized solution. However, in the transition to Healthcare 4.0, decentralization is of increasing importance. This concept is consistent with this development.

Future research also needs to further validate the use case and the proof of concept of VerifyMed. Before modifying and updating the proof of concept, a feasibility study with real users should be undertaken to validate the concept and explore the interface design. The feasibility study could also address the challenge of how a patient interprets the presented verification of experience and verification of the competence of a caregiver.

Authors' Contributions

AH contributed to the conceptualization, methodology, visualization, writing, and original draft preparation. JAHR contributed to the data curation, software, and visualization. KK contributed to the writing, reviewing, supervision, and conceptualization. DG contributed to the writing, reviewing, supervision, visualization, and conceptualization. AF was involved with the conceptualization, methodology, visualization, writing, reviewing, and supervision.

Conflicts of Interest

None declared.

References

1. Oppers E, Chikada K, Eich F, Imam P, Kiff J, Kisser M, et al. The financial impact of longevity risk. In: Global Financial Stability Report: The Quest for Lasting Stability. Washington, DC: International monetary fund; Apr 1, 2012:37.
2. Harris RE. Epidemiology of Chronic Disease: Global Perspectives: Global Perspectives. Burlington, Massachusetts, USA: Jones & Bartlett learning; 2019.

XSL•FO
RenderX

3. Bongaarts J. Global fertility and population trends. Semin Reprod Med 2015 Jan;33(1):5-10. [doi: 10.1055/s-0034-1395272] [Medline: 25565505]
4. Raikwar M, Gligoroski D, Kralevska K. SoK of used cryptography in blockchain. IEEE Access 2019;7:148550-148575. [doi: 10.1109/access.2019.2946983]
5. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. Bitcoin Org. 2019. URL: https://bitcoin.org/bitcoin.pdf [accessed 2021-07-13]
6. Mohanta BK, Panda SS, Jena D. An Overview of Smart Contract and Use Cases in Blockchain Technology. In: 9th International Conference on Computing, Communication and Networking Technologies. 2018 Jul 10 Presented at: ICCCNT'18; July 10-12, 2018; Bengaluru, India p. 1-4. [doi: 10.1109/icccnt.2018.8494045]
7. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-a scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
8. Buterin V. Blockchain Lab. 2014. URL: https://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf [accessed 2021-07-13]
9. Rolfe T. Turing Completeness and Smart Contract Security. Medium -Where Good Ideas Find You. 2019 Feb 1. URL: https://medium.com/kadena-io/turing-completeness-and-smart-contract-security-67e4c41704c [accessed 2021-05-08]
10. Kuo T, Rojas HZ, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 1;26(5):462-478 [FREE Full text] [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
11. Zhang M, Walker MA, White J, Schmidt DC, Lenz G. Metrics for Assessing Blockchain-based Healthcare Decentralized Apps. In: 19th International Conference on e-Health Networking, Applications and Services. 2017 Dec 18 Presented at: IEEE'17; October 12-15, 2017; Dalian, China. [doi: 10.1109/healthcom.2017.8210842]
12. Schmidt P. Blockcerts - An Open Infrastructure for Academic Credentials on the Blockchain. Medium - Where Good Ideas Find You. 2016 Oct 24. URL: https://tinyurl.com/6f3jeb5t [accessed 2021-07-13]
13. Baldi M, Chiaraluce F, Kodra K, Spalazzi L. Security Analysis of a Blockchain-based Protocol for the Certification of Academic Credentials. arXiv.org e-Print archive. 2019. URL: https://arxiv.org/abs/1910.04622 [accessed 2021-07-13]
14. A Digital Exchange for Professional Credentials Data. Hashed Health. 2020. URL: https://hashedhealth.com/blockchain-healthcare-solutions/professional-credentials-exchange/ [accessed 2021-05-01]
15. Axuall, Inc. and MetroHealth Announce Collaboration to Improve and Streamline Practitioner Credentialing. PR Newswire. 2020. URL: https://www.prnewswire.com/news-releases/axuall-inc-and-metrohealth-announce-collaboration-to-improve-and-streamline-practitioner-credentialing-300997828.html [accessed 2021-05-01]
16. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. Br Med J 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
17. Doos L, Packer C, Ward D, Simpson S, Stevens A. Past speculations of the future: a review of the methods used for forecasting emerging health technologies. BMJ Open 2016 Mar 10;6(3):e010479 [FREE Full text] [doi: 10.1136/bmjopen-2015-010479] [Medline: 26966060]
18. Pohl K. Requirements Engineering Fundamentals, Principles, and Techniques. Berlin/Heidelberg, Germany: Springer Publishing Company; Jul 23, 2010.
19. Hevner, March, Park, Ram. Design science in information systems research. MIS Quarterly 2004;28(1):75. [doi: 10.2307/25148625]
20. Thuemmler C. In: Thuemmler C, Bai C, editors. Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare. Swizerland: Springer International Publishing; 2017.
21. Lasi H, Fettke P, Kemper H, Feld T, Hoffmann M. Industry 4.0. Bus Inf Syst Eng 2014 Jun 19;6(4):239-242. [doi: 10.1007/s12599-014-0334-4]
22. Hermann M, Pentek T, Otto B. Design principles for Industrie 4.0 scenarios. In: 49th Hawaii International Conference on System Sciences. 2016 Mar 10 Presented at: HICSS'16; January 5-8, 2016; Koloa, HI, USA. [doi: 10.1109/hicss.2016.488]
23. Gregor S. A Theory of Theories in Information Systems. Information Systems Foundations: Building the Theoretical Base. 2002. URL: https://tinyurl.com/ua9ka254 [accessed 2021-07-13]
24. McKnight D, Chervany N. Trust and distrust definitions: one bite at a time. In: Trust in Cyber-societies Integrating the Human and Artificial Perspectives. Berlin, Heidelberg, Germany: Springer; Dec 20, 2001:27-54.
25. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. Acad Manag Rev 1998 Jul 1;23(3):393-404. [doi: 10.5465/Amr.1998.926617]
26. Coulter A. Patients' views of the good doctor. Br Med J 2002 Sep 28;325(7366):668-669 [FREE Full text] [doi: 10.1136/bmj.325.7366.668] [Medline: 12351342]
27. Mechanic D, Schlesinger M. The impact of managed care on patients' trust in medical care and their physicians. J Am Med Assoc 1996 Jun 5;275(21):1693-1697. [Medline: 8637148]
28. The duties of a doctor registered with the General Medical Council. General Medical Council. 1998. URL: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice/duties-of-a-doctor [accessed 2021-07-13]

XSL•FO
RenderX

29. The Nurse's Code: A Practical Approach to the Code of Professional Conduct for Nurses, Midwives and Health Visitors. London, UK: The British Psychological Society; 2018.
30. Croker JE, Swancutt DR, Roberts MJ, Abel GA, Roland M, Campbell JL. Factors affecting patients' trust and confidence in GPs: evidence from the English national GP patient survey. BMJ Open 2013 May 28;3(5):- [FREE Full text] [doi: 10.1136/bmjopen-2013-002762] [Medline: 23793686]
31. Sakallaris BR, Miller WL, Saper R, Kreitzer MJ, Jonas W. Meeting the challenge of a more person-centered future for US healthcare. Glob Adv Health Med 2016 Jan;5(1):51-60 [FREE Full text] [doi: 10.7453/gahmj.2015.085] [Medline: 26937314]
32. Leimeister JM, Ebner W, Krcmar H. Design, implementation, and evaluation of trust-supporting components in virtual communities for patients. J Manag Inf Syst 2014 Dec 8;21(4):101-131. [doi: 10.1080/07421222.2005.11045825]
33. O´neill O. Intelligent trust in a digital world. New Perspect Q 2017 Oct;34(4):27-31. [doi: 10.1111/npqu.12105]
34. Hall MA, Zheng B, Dugan E, Camacho F, Kidd KE, Mishra A, et al. Measuring patients' trust in their primary care providers. Med Care Res Rev 2002 Sep;59(3):293-318. [doi: 10.1177/1077558702059003004] [Medline: 12205830]
35. Pearson SD, Raeke LH. Patients' trust in physicians: many theories, few measures, and little data. J Gen Intern Med 2000 Jul;15(7):509-513 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.11002.x] [Medline: 10940139]
36. Frow P, McColl-Kennedy JR, Payne A. Co-creation practices: Their role in shaping a health care ecosystem. Ind Mark Manag 2016 Jul;56:24-39. [doi: 10.1016/j.indmarman.2016.03.007]
37. Montague EN, Kleiner BM, Winchester WW. Empirically understanding trust in medical technology. Int J Ind Ergon 2009 Jul;39(4):628-634. [doi: 10.1016/j.ergon.2009.01.004]
38. Technical Report. CompData. 2017. URL: https://tinyurl.com/ek2s8pwa [accessed 2021-07-13]
39. Najib M, Abdullah S, Narresh S, Juni MH. Brain-drain phenomenon among healthcare workers. Int J Public Health Clin Sci 2019 Jun 1;6(3):90-103. [doi: 10.32827/ijphcs.6.3.90]
40. Sabesan S, Kelly J, Evans R, Larkins S. A tele-oncology model replacing face-to-face specialist cancer care: perspectives of patients in North Queensland. J Telemed Telecare 2014 Jun;20(4):207-211. [doi: 10.1177/1357633X14529237] [Medline: 24643950]
41. Andreassen HK, Trondsen M, Kummervold PE, Gammon D, Hjortdahl P. Patients who use e-mediated communication with their doctor: new constructions of trust in the patient-doctor relationship. Qual Health Res 2006 Feb;16(2):238-248. [doi: 10.1177/1049732305284667] [Medline: 16394212]
42. Rensaa H. Verifymed - Application of Blockchain Technology to Improve Trust in Virtualized Healthcare Services. ResearchGate. Trondheim: Norwegian University of Science and Technology (NTNU); 2020 Jun 1. URL: https://www.researchgate.net/publication/349251550_VerifyMed_-_Application_of_blockchain_technology_to_improve_trust_in_virtualized_healthcare_services [accessed 2021-07-13]
43. Rensaa H, Gligoroski D, Kralevska K, Hasselgren A, Faxvaag A. Verifymed-a Blockchain Platform for Transparent Trust in Virtualized Healthcare: Proof-of-concept. In: Proceedings of the 2020 2nd International Electronics Communication Conference. 2020 Jul Presented at: IECC'20; July 8-10, 2020; Singapore p. 73-80. [doi: 10.1145/3409934.3409946]
44. Rensaa J. Transparent healthcare. GitHub. 2020. URL: https://github.com/jarensaa/transparent-healthcare [accessed 2021-05-01]
45. van Eecke P, Haie A. Practitioner’s corner · blockchain and the GDPR: the EU blockchain observatory report. Eur Data Prot L Rev 2018;4(4):531-534. [doi: 10.21552/edpl/2018/4/18]
46. Hasselgren A, Wan P, Horn M, Kralevska K, Gligoroski D, Faxvaag A. GDPR Compliance for Blockchain Applications in Healthcare. In: International Conference on Big Data, IoT and Blockchain. 2020 Oct Presented at: BIBC'20; October 24-25, 2020; Dubai, UAE URL: https://app.cristin.no/projects/show.jsf?id=2041854 [doi: 10.5121/csit.2020.101303]

Abbreviations

dApps: decentralized applications
GDPR: General Data Protection Regulation
HIPAA: American Health Insurance Portability and Accountability Act
PROM: patient-reported outcome measure

XSL•FO
RenderX

Edited by R Kukafka; submitted 04.03.21; peer-reviewed by R El-Gazzzar, D Cabrera, T Ueno, C Reis, TT Kuo; comments to author 26.04.21; revised version received 19.05.21; accepted 14.06.21; published 30.07.21.

Please cite as:
Hasselgren A, Hanssen Rensaa JA, Kralevska K, Gligoroski D, Faxvaag A
Blockchain for Increased Trust in Virtual Health Care: Proof-of-Concept Study
J Med Internet Res 2021;23(7):e28496
URL: https://www.jmir.org/2021/7/e28496
doi:10.2196/28496
PMID:34328437

©Anton Hasselgren, Jens-Andreas Hanssen Rensaa, Katina Kralevska, Danilo Gligoroski, Arild Faxvaag. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Policy Interventions, Development Trends, and Service Innovations of Internet Hospitals in China: Documentary Analysis and Qualitative Interview Study

Yunfeng Lai[1*], PhD; Shengqi Chen[1*], PhD; Meng Li[1], PhD; Carolina Oi Lam Ung[1], PhD; Hao Hu[1], PhD

State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macau, China

[*]these authors contributed equally

Corresponding Author:
Hao Hu, PhD
State Key Laboratory of Quality Research in Chinese Medicine
Institute of Chinese Medical Sciences
University of Macau
Room 2057, Research Building N22
Avenida da Universidade, Taipa, Macau
Macau, 999078
China
Phone: 853 88228538
Email: haohu@um.edu.mo

Abstract

Background: Internet hospitals have been encouraged by the Chinese government to develop an innovative medical service model that mainly uses new internet-based technologies to increase access to health care and improve the quality and efficiency of health care delivery. However, the academic exploration of the institutional and sectoral development of internet hospitals in China is scarce in the existing literature.

Objective: This study aimed to investigate the policy interventions, development trends, and service innovations of internet hospitals in China. It is expected that the findings from this study will contribute to the further innovation of internet hospitals in China and provide references for the international development of internet hospitals for personalized digital health and patient-centric services.

Methods: This study analyzed official policies related to internet hospitals that were implemented by the government in China since 2005. The data of formally approved internet hospitals were collected from official websites to analyze development trends. In-depth semistructured interviews were conducted with 58 key stakeholders who represented comprehensive viewpoints about the service innovations of internet hospitals between March and November 2019.

Results: In total, 25 policies that promoted the development of internet hospitals in China were identified. These policies encompassed informatization infrastructure construction, medical resource integration, development model design, service model design, and payment model design. Of the 268 internet hospitals that had received an official license from the government, 153 public internet hospitals had been built mainly by medical institutions. Public tertiary hospitals were the main actors in founding internet hospitals that were created to provide services that targeted patients with common diseases or chronic diseases or patients living in remote and rural areas. Promoting convenient access to high-quality medical resources and saving patients' and their families' time were the key values of internet hospitals.

Conclusions: The policy interventions strongly promoted the development of internet hospitals in China. Public tertiary hospitals led the development of internet hospitals. However, internet hospitals in China have mainly played roles that are complementary to those of physical medical institutions. The service model of internet hospitals needs more distinguished innovations to provide personalized digital health and patient-centric services.

(J Med Internet Res 2021;23(7):e22330) doi:10.2196/22330

KEYWORDS

internet hospital; health policy; medical service; public hospital; digital health; China

XSL•FO
RenderX

Introduction

A total of 10 years have passed since the launch of new health reform in China. Substantial progress in providing equal access to health care and improving financial risk protection has been achieved. However, there are still gaps in medical quality, the control of noncommunicable disease, the control of health expenditures, public satisfaction, and the distribution of medical resources [1]. Indeed, due to the vast population and uneven distribution of medical resources in China, the use of medical resources is contentious [2]. In order to promote China's health care, there is a need to further reform the medical payment system, the hospital governance structure, primary health care institutions' capacity-building methods, and health education and a need to make full use of digital medical technology [3-5]. The expected outcomes are the promotion of the construction of health-oriented medical alliances [6-8], improvements in medical service quality through medical insurance payments, and the acceleration of medical service innovation [9]. As such, internet hospitals have been encouraged in recent years by the Chinese government.

In order to develop internet hospitals effectively, the Chinese government issued various relevant policy interventions in the past 15 years. The government aimed to develop an innovative medical service model that mainly used new internet-based technologies to increase access to health care and improve the quality and efficiency of health care delivery. Notably, technological innovations were used to improve health care use across hospitals in China [10]. As an emerging innovation in China, internet hospitals are expected to improve the availability of medical resources and decrease the costs of distant medical services [11].

Telehealth (or telemedicine) is 1 kind of worldwide medical service model that is similar to the internet hospital model in China [12]. In advanced countries like the European Union, the United States, and Japan, telehealth has been studied and practiced for many years, mainly in fields such as ophthalmology, cardiovascular, and dermatology [13]. However, in terms of service scope and means, the internet hospital model in China is more advanced than the traditional telehealth model. In theory, internet hospitals in China are internet-based medical platforms that medical institutions use to directly provide medical services to patients by using information technology to extend medical resources from hospitals to the internet and carry out web-based medical care and health services that can be received by patients in a home setting [14]. The internet hospital model is a new format that is derived from the "Internet + Health Care" format, which is an extension of the telemedicine and traditional hospital formats. By using the medical resources of physical hospitals and internet technology, internet hospitals provide patients with closed-loop medical services (ie, web-based and offline medical services, including web-based registration, web-based medical consultation, web-based drug supply and support, and web-based payment services), so that patients can use the more convenient medical services of physical hospitals and optimize and match existing medical needs and health care resources [15]. It is believed that the scope of internet hospitals not only encompasses telemedicine but also covers electronic prescriptions, medical insurance, commercial health insurance, health management, hospital operation, and hospital logistics. The internet hospital model is expected to become an innovative medical service model that is different from traditional hospital models [16]. In particular, during the COVID-19 pandemic, internet hospitals in China have become important components of infection prevention and control measures by offering essential medical support to the public despite the strict social distancing requirements, promoting self-care and self-protection, and facilitating epidemiological screening [17,18].

At present, the development of internet hospitals is far from mature and is limited by many factors, such as the scarcity of web-based doctors and the unavailability of medical insurance coverage [19]. Existing literature on internet hospitals mainly focuses on informatization infrastructure construction and development model design [20]. Consequently, while internet hospitals have been encouraged by the Chinese government to optimize health care resource use, the academic exploration of the institutional and sectoral development of internet hospitals in China remains scarce in existing literature. Thus, this study aimed to investigate the policy interventions, development trends, and service innovations of internet hospitals in China. It is expected that the findings from this study will contribute to the further innovation of internet hospitals in China and provide references for the international development of internet hospitals for personalized digital health and patient-centric services.

Methods

Research Design

This study had a qualitative research design and collected data via the following three methods: (1) policy file collection and analysis; (2) the documentary analysis of internet hospital approval; and (3) qualitative interviews with key stakeholders.

Policy File Collection and Analysis

To understand the policy interventions on internet hospital development in China, we collected and analyzed the relevant policy files of internet hospitals. We searched for these files on the websites of relevant government departments, including the websites of the State Council, the National Health Commission, the National Administration of Traditional Chinese Medicine, the National Health Security Administration, and the National Medical Product Administration. These websites were searched for relevant policies and regulations describing "Internet + Health Care" and internet hospitals. We used the following search terms: *Internet hospital*, *Internet + health*, *Internet + healthcare*, *Internet medical consultation*, *telemedicine*, *Internet + medical insurance service*, *Internet drug business*, *healthcare big data*, *health information*, *Healthy China 2030*, *tiered healthcare delivery*, *medical treatment combination* (*yi lian ti*), and *medical alliance* (*yi gong ti*). We also used combinations of these terms. The date of the last search was March 10, 2020.

In terms of the inclusion and exclusion criteria, this study only included formal and legal policy documents from official authorities. These selected policy documents were reviewed

XSL•FO
RenderX

based on their objectives and their relevance to supporting internet hospitals by (1) providing a legal framework for practice, (2) identifying a development plan, (3) formulating specific development actions, (4) formulating management actions, and (5) formulating security measures.

Documentary Analysis of Internet Hospital Approvals

To explore the development trends of internet hospitals in China, we conducted a documentary analysis of the official approvals of internet hospitals. With regard to the data source, the data of formally approved internet hospitals that were published up to April 15, 2020, were directly collected from the official websites of provincial health commissions in China—the competent authorities that approve the establishment of internet hospitals in their provinces. The names, locations, owners, and approval times of internet hospitals were directly searched on and extracted from the official websites of the provincial health commissions by our research term. After collecting all of the data, we conducted descriptive statistical analyses to analyze the number, geographical distribution, approval time, ownership, and founding institutions (medical institution vs company) of established internet hospitals.

Qualitative Interviews With Key Stakeholders

To investigate key stakeholders' perceptions about the service innovations of internet hospitals, we used a qualitative interviewing method. Between April and November 2019, semistructured interviews were conducted with a wide range of key stakeholders in the internet hospital sector, including policy makers, not-for-profit professional societies, hospital administrators, doctors, managers of pharmaceutical companies, managers of internet companies, information technology engineers, academic researchers, and financial investors.

Purposive sampling was used to select the interviewees. A total of 58 key stakeholders took part in the interviews, and their characteristics are presented in Table 1. There were 44 males and 14 females. Interviewees were from 10 regions and included 26 stakeholders from Beijing, 16 from Guangdong, 6 from Sichuan, 4 from Shanghai, 2 from Zhejiang, 1 from Chongqing, 1 from Tianjin, 1 from Jiangsu, and 1 from Hubei.

Table 1. Characteristics of interviewees (N=58).

Characteristics	Interviewees, n
Policy makers	9
Not-for-profit professional societies	5
Hospital management staffs	8
Doctors	5
Pharmaceutical company managers	12
Internet company managers	9
Information technology engineers	2
Academic researchers	7
Financial investors	1

We approached the interviewees in advance to obtain informed consent. In total, 45 interviews were conducted face-to-face, and 13 interviews were conducted via telephone. The interviewees were asked questions that primarily focused on the policy interventions, development trends, and service innovations of internet hospitals. We recorded the audio of all interviews, which were transcribed verbatim with the interviewees' consent. The average interview time was about 50 minutes.

Thematic analysis was used for the data analysis of qualitative interviews. First, two researchers conducted the thematic analysis separately. Second, they met with each other to identify similarities and the differences by comparing the analysis results. Third, another two researchers analyzed the differences to conduct a triangular test. Finally, the final results of the qualitative investigations were reviewed by all researchers together.

Data Availability

The data sets generated for this study are available from the corresponding authors upon reasonable request.

Ethics Statement

The study design was approved by the ethics committee of the University of Macau (approval number: BSERE20-APP004-ICMS).

Results

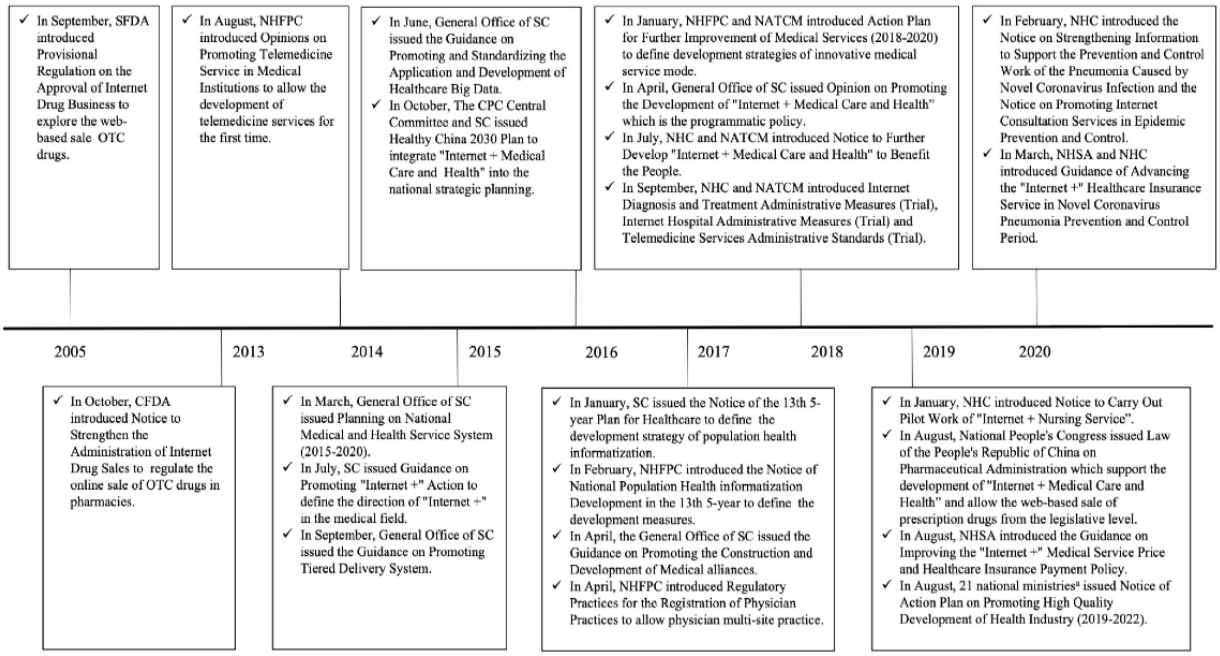
Policy Interventions of Internet Hospitals in China

Summary of Policy Interventions

In total, 25 policy files were identified in this study [21-43], including 1 law and 24 regulations. The details of these 25 policy documents are summarized in Multimedia Appendix 1. Corresponding with the official announcement timing of policy interventions, the policy developments of internet hospitals in China can be classified into two phases—phase 1 (2005-2017) and phase 2 (2018 onward; Figure 1).

XSL•FO
RenderX

Figure 1. The main policy interventions in the history of internet hospitals in China. These policy interventions were created by the following 21 national ministries: the National Development and Reform Commission, the Ministry of Education, the Ministry of Science and Technology, the Ministry of Industry and Information Technology, the Ministry of Civil Affairs, the Ministry of Finance, the Ministry of Human Resources and Social Security, the Ministry of Natural Resources, the Ministry of Ecology and Environment, the Ministry of Housing and Urban-Rural Development, the Ministry of Commerce, the Ministry of Culture and Tourism, the NHC, The People's Bank of China, State Taxation Administration, the State Administration for Market Regulation, the General Administration of Sport of China, the China Banking and Insurance Regulatory Commission, the National Healthcare Security Administration, the NATCM, and the National Medical Products Administration. CFDA: China Food and Drug Administration; CPC: Communist Party of China; NATCM: National Administration of Traditional Chinese Medicine; NHC: National Health Commission; NHFPC: National Health and Family Planning Commission; NHSA: National Health Security Administration; OTC: over-the-counter; SC: State Council; SFDA: State Food and Drug Administration.



Phase 1 (2005-2017)

In 2005, the State Food and Drug Administration allowed the web-based sales of over-the-counter drugs for the first time, which prompted the exploration of innovative drug services on internet [26]. In 2014, the National Health and Family Planning Commission allowed telemedicine service providers to create medical care services. In 2015, the State Council integrated the “Internet + Health Care” format into the national medical and health service system to promote health information services and smart medical services and benefit the whole population via information technology. Moreover, the State Council issued guidance on building a tiered health care delivery system to reform the existing hospital-centric model by using information technology and big data. In 2016, the country’s long-term health sector strategy—“Health China 2030”—was announced by President XI Jinping. This strategy involved integrating the “Internet + Health Care” format into the national strategic plans for health. Similarly, the use of health care big data in in key informatization infrastructure construction has also been the focus of the State Council. In 2017, the State Council promoted the development of medical institution alliances to operationalize the tiered health care delivery approach, which can strengthen the support and training that higher medical institutions provide to primary medical institutions through digital technology. The establishment of a population health information system was listed as a key project by the State Council and the National Health and Family Planning Commission. Furthermore, the National Health and Family Planning Commission allowed physicians to conduct multisite practices and gave official permission to physicians to work for internet-based medical institutions like internet hospitals.

Phase 2 (2018 onward)

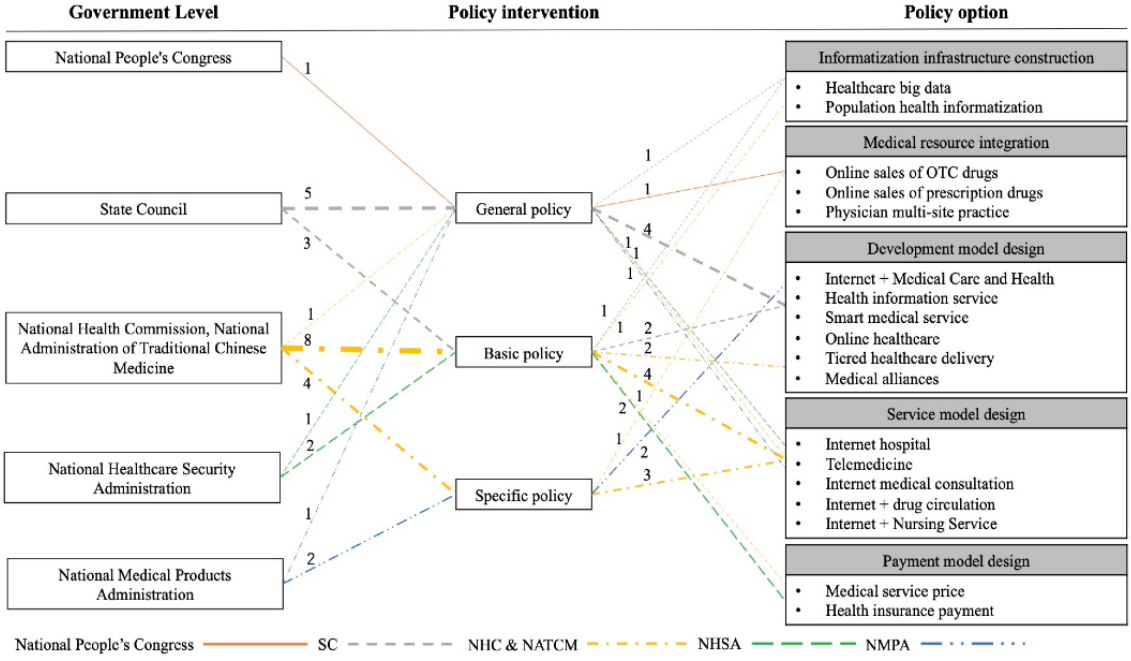
In 2018, a milestone was achieved in the development of internet hospitals. The State Council issued a programmatic policy that nationally recognized internet hospitals for the first time. Subsequently, the National Health Commission and the National Administration of Traditional Chinese Medicine issued the “Internet Medical Consultation Administrative Measures (Trial),” “Telemedicine Services Administrative Standards (Trial),” and “Internet Hospital Administrative Measures (Trial)” policies. These polices defined the access, practice, and supervision rules and measures of internet hospitals. In 2019, the National People's Congress issued the Law of the People's Republic of China on Pharmaceutical Administration, which clearly supported the development of “Internet + Health Care” hospitals and allowed the web-based sale of prescription drugs at the legislative level. The National Health Security Administration issued guidance on “Internet +” medical service prices and medical insurance payment policies in an attempt to support the medical insurance coverage of internet hospitals.

In general, with regard to policy focus, phase 1 emphasized informatization infrastructure construction, medical resource integration, and development model design; phase 2 prioritized the transformation of resources into effective services through systemic service model design and payment model design. A

XSL•FO
RenderX

detailed analysis of policy makers, policy levels, and policy options are summarized in Figure 2.

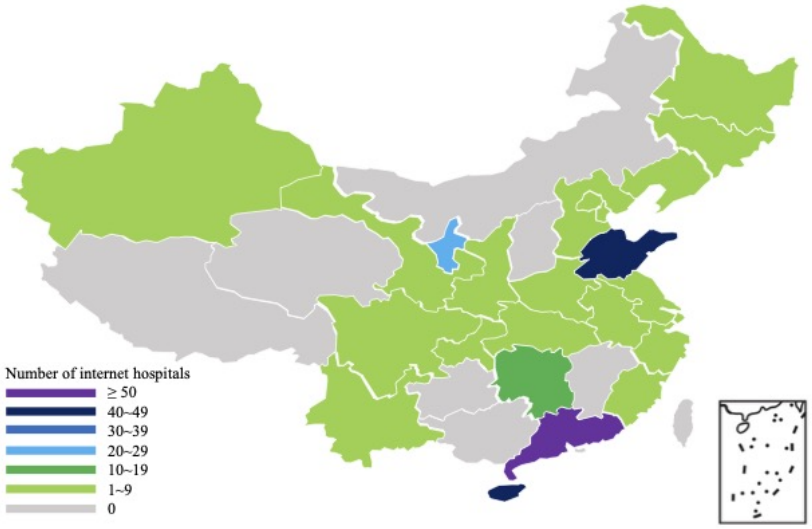
Figure 2. The evolution of policy options for internet hospitals in China. The numbers represent the number of policy documents. NATCM: National Administration of Traditional Chinese Medicine; NHC: National Health Commission; NHSA: National Healthcare Security Administration; NMPA: National Medical Products Administration; OTC: over-the-counter; SC: State Council.



Development Trends of Internet Hospitals in China

Until April 15, 2020, a total of 268 internet hospitals received official licenses from the government that were distributed in 24 provinces (municipalities) in China (Figure 3). Guangdong Province, Shandong Province, and Hainan Province approved more than 40 internet hospitals and thus played leading roles in the development of internet hospitals.

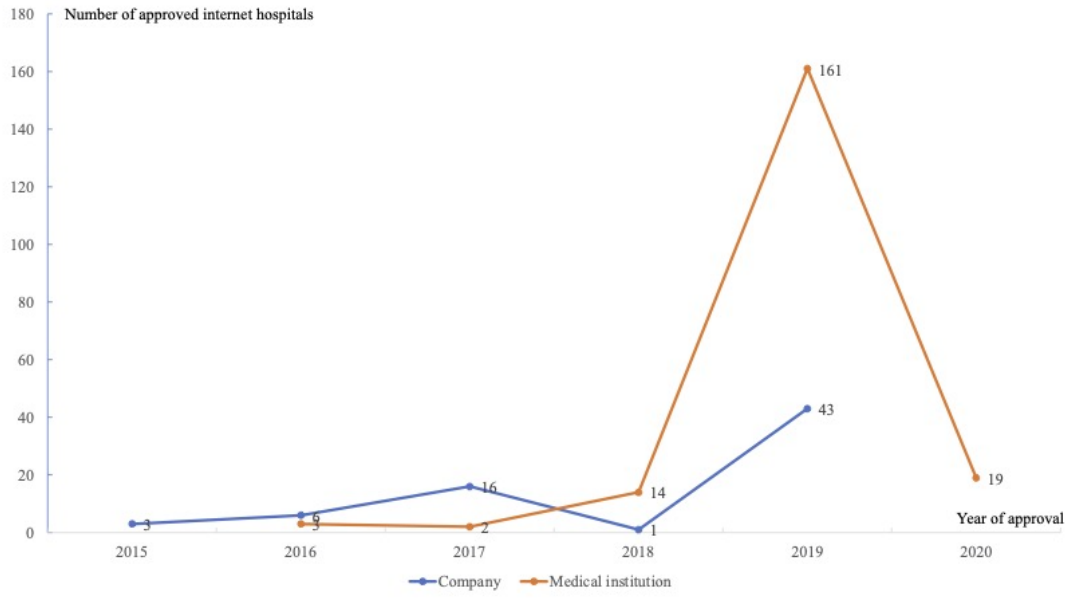
Figure 3. The geographical distribution of internet hospitals in China.



XSL•FO
RenderX

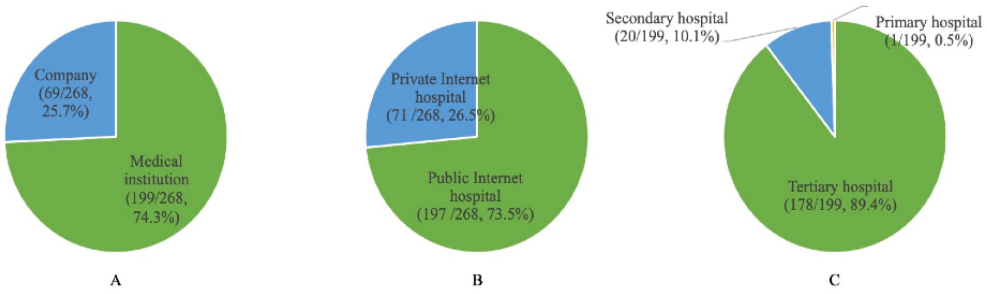
As shown in Figure 4, the establishment of internet hospitals increased quickly since 2019. This was the result of the new policy support released in 2018. In 2019, a total of 204 internet hospitals were approved (including 161 internet hospitals founded by medical institutions and 43 internet hospitals founded by companies), accounting for 76.1% (204/268) of internet hospitals in China.

Figure 4. Approval of internet hospitals in China. No internet hospitals that were founded by companies were approved between January 1, 2020, and April 15, 2020.



Of the 268 approved internet hospitals, 199 (74.3%) internet hospitals were established by medical institutions, which is much higher than the number of internet hospitals founded by companies such as internet companies and risk investment companies (Figure 5). This implies that medial institutions have more resources and the capacity to develop internet hospitals. With regard to the ownership of internet hospitals, there were 197 public internet hospitals that were mainly established by medical institutions and 71 private internet hospitals (Figure 5).

Figure 5. Proportions of internet hospital types. A: The proportions of internet hospitals developed by medical institutions and companies. B: The proportions of public internet hospitals and private internet hospitals. C: The proportions of internet hospitals established by different levels of medical institutions.



In total, 199 internet hospitals were founded by medical institutions; the number of internet hospitals established by tertiary hospitals, secondary hospitals, and primary hospitals was 178 (89.4%), 20 (10.1%), and 1 (0.5%), respectively (Figure 5). These data show that tertiary hospitals, which have the absolute advantage in terms of medical resources, have more motivations and the ability to establish internet hospitals. Furthermore, 36.4% (75/206) of tertiary hospitals in Guangdong Province founded their internet hospitals, 23.1% (3/13) of tertiary hospitals in Ningxia Province founded their internet hospitals, and 22.7% (41/181) of tertiary hospitals in Shandong Province founded their internet hospitals (Table 2).

XSL•FO
RenderX

Table 2. Development of internet hospitals established by different levels of medical institutions.

Location	Tertiary hospitals (N=178), n	Secondary hospitals (N=20), n	Primary hospitals (N=1), n	Internet hospitals in tertiary hospitals[a], n (%)
Guangdong	75	6	0	75 (36.4)
Shandong	41	7	0	41 (22.7)
Hunan	10	0	0	10 (12.7)
Jiangsu	8	0	0	8 (5)
Shanghai	4	1	0	4 (8.5)
Yunnan	2	3	0	2 (2.9)
Sichuan	5	0	0	5 (2.5)
Fujian	5	0	0	5 (6.5)
Gansu	5	0	0	5 (13.5)
Auhui	5	0	0	5 (7.4)
Zhejiang	3	0	1	3 (2.2)
Liaoning	3	0	0	3 (2.2)
Ningxia	3	0	0	3 (23.1)
Heilongjiang	2	0	0	2 (2.1)
Tianjin	2	0	0	2 (4.7)
Shanxi	1	1	0	1 (1.6)
Xinjiang	0	1	0	0 (0)
Hebei	1	0	0	1 (1.4)
Beijing	1	0	0	1 (1)
Hubei	1	0	0	1 (0.8)
Henan	1	0	0	1 (4.6)
Jinlin	0	1	0	0 (0)

[a]The total proportion of internet hospitals in tertiary hospitals was 8.8% (178/2015).

Service Innovations of Internet Hospitals in China

Unlike traditional medical institutions, internet hospitals in China have tried to implement service innovations, which can be categorized based on the following three aspects: the target patients of internet hospitals, value offerings of internet hospitals, and the services provided by internet hospitals.

Target Patients of Internet Hospitals

With regard to the target patients of internet hospitals, there are two main types of patients—patients with common diseases or chronic diseases and patients in remote and rural areas.

Patients With Common Diseases or Chronic Diseases

At present, internet hospitals are not allowed to provide medical services to first-diagnosis patients due to certain policies. Internet hospitals were mainly designed to provide medical services to countercheck patients with common diseases or chronic diseases. An interviewee said:

> *At present, the Internet hospital is still in the exploratory development stage. For the sake of people's life, health and safety, the first step is to open only to the patients with common diseases or chronic diseases. When the conditions are mature, we will gradually open to the first-diagnosis patients.* [Interviewee #2, policy maker at the National Health Commission]

Patients in Remote and Rural Areas

Due to objective conditions such as economic underdevelopment and a lack of medical resources, problems such as an insufficient number of primary physicians and the unmet health care needs of primary patients in remote and rural areas are very prominent. Internet hospitals were thus positioned to meet the unmet medical service needs of patients in remote and rural areas. This was described by two interviewees, as follows:

> *For example, some stroke patients from rural areas do not know how to recover at home after discharge. But the village doctors there are unable to provide professional advice either. In such kind of situation, it needs specialist doctor (of Internet hospital) to tell the patients what to do and how to do.* [Interviewee #5, manager of an internet company]

> *Every village doctor in Guangdong Province manages more than 1,000 patients on average, even more than 5,000 patients. They can not meet the needs of health screening, health follow-up, and clinical instruction*

XSL•FO
RenderX

of so many patients. [Interviewee #15, doctor at a public hospital]

The Key Values of Internet Hospitals

Unlike the traditional medical institutions in China, internet hospitals might offer the following three distinctive values: saving patients' and their families' time, providing more convenient access to high-quality medical resources, and expanding from treatment-oriented services to complete health management.

Saving Patients' and Their Families' Time

One of the main problems that result in poor medical experiences in China is the time cost of medical services at traditional hospitals, including long registration times, long waiting times, short medical consultation times, and long times for taking medicine. With the help of internet hospitals, these problems could be greatly improved. This was expressed by two interviewees, as follows:

> *80% of the patients in Southern Hospital are non-local. It is very troublesome for countercheck patients every time. For example, a patient with asthma in Hunan Province needs to spend three days per month to take a high-speed train to come to the hospital for medical consultation and medicine. The emergence of Internet hospital can solve this problem well. He only needs to communicate with the doctor through pictures and texts or video, and the doctor can send the medicine by express to him after diagnosis, which can save a lot of time for patient.* [Interviewee #29, management staff member of a public hospital]

> *The most obvious advantage of Internet hospital is the convenience of registration, waiting, medical consultation and medicine taken, so as to avoid wasting time and energy.* [Interviewee #32, policy maker at the City Health Commission]

More Convenient Access to High-Quality Medical Resources

Internet hospitals overcame geographical restrictions to provide medical services to grassroots patients directly. In theory, with the help of internet hospitals, patients can obtain medical services from any hospital, including exceptionally scarce, high-quality medical resources at high-level medical institutions. This was described by two interviewees, as follows:

> *Internet hospital is mainly to solve the medical problems and alleviate the low competence challenge of doctors in primary hospitals.* [Interviewee #5, manager of an internet company]

> *Internet hospital has two core functions: one is to solve the shortage of human resources in primary hospitals; the other is to solve the uneven distribution of medical resources. For example, in Guangdong Province, 80% of medical resources are concentrated in Guangzhou City, but 90% of patients are at other cities.* [Interviewee #15, doctor of a public hospital]

Service Expansion From Treatment-Oriented Services to Complete Health Management Services

Unlike the traditional medical institutions that focused primarily on medical treatment, internet hospitals were expected to provide complete health management services for patients and ordinary persons. This was explained by two interviewees, as follows:

> *Large hospitals only solve the short-term problem of patients' acute attack. But patients' later rehabilitation is more important, including patients' health recording and health education, which is the biggest problem of medical system. We should consider how to make good use of Internet hospital for health management at the grass-roots level.* [Interviewee #15, doctor of a public hospital]

> *The development of Internet hospitals should insist on people-centered and expand medical services from treatment-centered to health-centered.* [Interviewee #2, policy maker at the National Health Commission]

Services Provided by Internet Hospitals

According to the service process, the main services provided by internet hospitals can be categorized as web-based registration services, web-based medical consultation services, web-based drug supply services, and web-based payment services.

Web-Based Registration Service

Internet hospitals provide web-based registration services for both outpatients of physical medical institutions and patients attending web-based medical consultations. The web-based outpatient registration of physical medical institutions covered all medical departments. Comparatively, web-based registration for web-based medical consultations was only open to patients with common diseases or chronic diseases. The was described by an interviewee, as follows:

> *In the past, online registration was only used for outpatients of physical medical institution, for patient convenience. Now it (online registration service) has become the entrance of Internet hospitals, having the function of triage patients.* [Interviewee #40, management staff member of a public hospital]

Web-Based Medical Consultation Service

According to different service models, consultation services can be divided into three types—image-text medical consultations, telephone medical consultations, and video medical consultations. These are different in terms of service objects, service content, and prescriptions.

With regard to image-text medical consultations, patients need to fill in the application form for medical consultations, which includes patients' health-related information (eg, condition description, past medical history, allergies, family genetic diseases, marriage, and birth status, and personal habits). During this time, patients can provide relevant health-related information by sending images. For counterchecked patients, physicians could make web-based diagnoses and order a prescription after reviewing specific health-related information.

XSL•FO
RenderX

For first-diagnosis patients, physicians could only provide health consulting services that did not involve disease diagnosis and prescription.

With regard to telephone medical consultations, patients first need to fill in the application form for medical consultations first before scheduling a time with the doctor for telephone medical consultation. However, doctors can only provide health consultations; they cannot provide disease diagnoses and prescriptions.

Video medical consultations could be used for counterchecking patients. After completing the application form, video medical consultations are carried out at the time that was agreed upon by both the patients and the physicians. The physician can then make a web-based diagnosis and order a prescription after reviewing specific health-related information. An interviewee presented his ideas about the three types of web-based consultation services, as follows:

> *Image-text medical consultation, telephone medical consultation, and video medical consultation are three common service models at present. In fact, most patients choose telephone consultation because it is convenient and not as troublesome as text description. However, from the perspective of hospital or doctor, they prefer image-text and video consultation, it's more objective and accurate, and is easy to keep records.* [Interviewee #34, management staff member of a public hospital]

Web-Based Drug Supply Service

After making a web-based disease diagnosis and offering a prescription, internet hospitals can provide web-based drug services. Internet hospitals can either send drugs to patients directly or transfer prescriptions via web-based platforms to designated community pharmacies that were near patients. Two interviewees described this service, as follows:

> *At present, online prescriptions only flow to the pharmacy of the hospital,.... Patients can go to get the drugs directly or send drugs by express to their home.* [Interviewee #29, management staff member of a public hospital]

> *For example, in Wuzhou City, the city platform of prescription circulation has been established. Community pharmacies joining the platform can accept online prescription orders. Patients can pick up drugs at the community pharmacy around their home or receive drugs by express delivery.* [Interviewee #51, academic researcher at a university]

Web-Based Payment Service

Internet hospitals can provide web-based payment services for expenses related to registration, medical consultations, and drugs. Due to the differences in the progress of medical insurance reform in different regions, the web-based settlement of medical insurance had only been realized in a few regions of China. However, patients need to pay their expenses if there was no web-based medical insurance settlement. This was explained by two interviewees, as follows:

> *Medical insurance is often the last one that gets involved. As the payer, we must be cautious. Moreover, the medical insurance policy of Internet hospital has not been fully developed in the country, mainly for pilots. Only a few provinces, such as Shandong Province and Fujian Province, have issued specific policies, such as reimbursement items and price.* [Interviewee #22, policy maker at the City Health Commission]

> *If the Internet hospital is not connected to medical insurance, all the expenses of patients in the Internet hospital need to be paid by themselves and cannot be reimbursed.* [Interviewee #17, internet hospital manager of an internet company]

Discussion

Principal Findings

This study reported a detailed overview of the policy interventions, development trends, and service innovations of internet hospitals in China. The understanding of these internet hospital aspects is lacking in existing literature. Our findings contribute to the understanding of policy development and implementation in the internet hospital sector. Based on our findings, there are some points that are worthy of further discussion.

First, after collating and summarizing new policies for internet hospitals, it became apparent that the multiple policy interventions introduced in the internet hospital sector over the past 15 years reflected policy makers' determination in firmly supporting the development of internet hospitals. Moreover, these continuous policy interventions gave stakeholders the confidence to found internet hospitals and integrate them into the entirety of the health system. This is particularly important for resource-limited countries such as China, where there is a rising demand for health care but an uneven distribution of medical resources [44]. Although the Chinese government has been promoting the development of a tiered health care delivery system that is anchored in primary health care, the actual implementation of this system deviated substantially from the ideal model [45,46]. Therefore, based on the rapid development of internet hospitals in China, the medical care and health services of medical institutions at all levels can be coordinated and integrated, which can lead to the improvement of medical alliances and tiered health care delivery [47]. It would be beneficial to solve the problem of the uneven distribution of medical resources, innovate health care, and improve the efficiency of medical services.

Based on the evolution of policy interventions, internet hospitals feature the concept of "experimentation precedes popularization" [48]. The first step in achieving these goals was to identify potential problems by piloting web-based sales of over-the-counter drugs and telemedicine services. In the second step, informatization infrastructure construction, medical resource integration, and development model design were carried out based on the pilot tests. In the final step, the service model and payment model were introduced to strengthen the development of internet hospitals. All of these processes show

XSL•FO
RenderX

that the Chinese government, which is cautious about the development of internet hospitals because they are a type of unknown medical institution, has taken a step-by-step approach to design and implement policies related to these five processes. These policy designs provide valuable references to other countries' policy development processes for internet hospitals.

Second, this study showed that public tertiary hospitals played a more leading role in establishing internet hospitals compared to the role of companies, such as large internet companies. This is consistent with previous research findings [49]. Since medical resources are mainly found at tertiary hospitals, they have distinct advantages in terms of medical equipment, finances, physicians, and patient sources [50]. Furthermore, public tertiary hospitals employ most of the high-level physicians and own most of the medical facilities in China [51]. Therefore, most of the existing internet hospitals were founded by public tertiary hospitals. Although internet companies have inherent advantages in terms of information technology and their ability to operate the internet, which are lacking in most traditional hospitals, they still cannot directly compete with public tertiary hospitals in China. In practice, internet companies have to cooperate with medical institutions to enter the sector of internet hospitals [52]. All of these findings indicate that high-quality medical resources like physicians and medical equipment are the key factors that shape the development of internet hospitals.

Third, the services provided by internet hospitals are particularly valuable to patients with common diseases or chronic diseases and patients in remote and rural areas. First-diagnosis patients are still unable to access the services provided by internet hospitals. In theory, the most common barriers that stop patients from obtaining satisfactory medical services are cost, access, skill, and time [53]. High-quality medical resources are mainly found in public tertiary hospitals [54]. However, due to financial and professional reasons, skilled physicians are unwilling to work in communities or remote and rural areas [55]. Additionally, many primary patients are reluctant to go to primary hospitals due to a lack of confidence in health professionals' skills and the quality of health care provided [56]. With the operation model that is currently in place, internet hospitals can reinforce the interactions between physicians and patients without the challenges of geographical limitations; increase the accessibility of high-quality medical resources in remote and rural areas; and dramatically reduce the indirect costs of medical care for patients, especially the expenses and time associated with travel [57]. As a result, regardless of the service type or service content, internet hospitals can at least play a role that is complementary to that of physical hospitals. Furthermore, due to innovations such as web-based registration services, web-based medical consultations, web-based drug supply and support services, and web-based payment services, internet hospitals can provide patients with more convenient methods for accessing high-quality medical resources and save patients' and their families' time. However, internet hospitals do not have direct cost and skill advantages.

In general, this study found that the government has provided a lot of space for the development of internet hospitals in China, but the value of internet hospital services has not been fully realized in operation. Thus, we propose the following suggestions for the further development of internet hospitals. First, although general policy interventions regarding medical insurance for internet hospital services have been issued, there are no specific implementation instructions at the national level. In particular, there is a need to introduce specific medical insurance measures at the national level as implementation guidance for local governments. Second, to motivate physicians at public tertiary hospitals to support the platform provided by internet hospitals, policies for encouraging physicians at public tertiary hospitals to join internet hospitals and providing reasonable incentives are needed. Third, policies for encouraging the sharing of medical facilities and equipment among internet hospitals and physical medial institutions are necessary to further the service innovations of internet hospitals.

Limitations

To the best of our knowledge, this is the first study that has collated and evaluated multiple policy interventions that were designed to influence internet hospital development in China. There are several research limitations that can be addressed in future studies. First, in this study, we did not collect data from patients. Further studies can focus on patients' realistic experiences with internet hospitals. Second, public hospital and internet companies have different strategies for developing internet hospitals, and these require further comparison and analysis.

Conclusions

Policy interventions regarding informatization infrastructure construction, medical resource integration, development model design, service model design, and payment model design have significantly promoted the development of internet hospitals in China. Further, public tertiary hospitals play a more leading role in founding internet hospitals compared to the role of internet companies. The service innovations of internet hospitals need to be further advanced with the support of corresponding policy interventions.

Acknowledgments

This study was partially supported by funding from the University of Macau (grant MYRG2019-00025-ICMS). The authors would like to thank all of the interviewees for their participation in this study.

XSL•FO
RenderX

Authors' Contributions

YL, SC, COLU, and HH conceptualized and designed the study. All authors collected the data and conducted the data analysis and data interpretation. YL, SC, COLU, and HH drafted the manuscript. ML participated in manuscript revision. All authors reviewed and approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the 25 policy documents.
[DOCX File , 27 KB - jmir_v23i7e22330_app1.docx]

References

1. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet 2019 Sep 28;394(10204):1192-1204. [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
2. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet 2008 Nov 15;372(9651):1774-1781. [doi: 10.1016/S0140-6736(08)61363-X] [Medline: 18930528]
3. The Lancet. Medical education reform in China. Lancet 2017 Jul 22;390(10092):334. [doi: 10.1016/S0140-6736(17)31921-9] [Medline: 28745589]
4. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan 07;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
5. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019 Mar 11;25(3):433-438. [doi: 10.1038/s41591-018-0335-9] [Medline: 30742121]
6. Shen L, Huang C, Li L, Sun G. Exploration and practice of county medical community model — Taking Tianchang city, Anhui province as an example. Health Econ Res 2018 Dec;12(380):7-11. [doi: 10.14055/j.cnki.33-1056/f.2018.12.002]
7. Liang S, He L, Song S, Jin Y, Yuan B, Meng Q. The development and practice of integrated healthcare in China. Chinese Journal of Health Policy 2016 May;9(5):42-48. [doi: 10.3969/j.issn.1674-2982.2016.05.008]
8. Lin L, Dai T, Zhu X. Analysis on the practice of health care alliance reform in Tianchang county of Anhui. Chinese Health Economics 2017 Apr;36(4):74-77. [doi: 10.7664/CHE20170419]
9. Wang L, Suo S, Li J, Hu Y, Li P, Wang Y, et al. An investigation into traditional Chinese medicine hospitals in China: Development trend and medical service innovation. Int J Health Policy Manag 2017 Jan 01;6(1):19-25 [FREE Full text] [doi: 10.15171/ijhpm.2016.72] [Medline: 28005539]
10. Tu J, Wang C, Wu S. Using technological innovation to improve health care utilization in China's hospitals: the emerging 'online' health service delivery. J Asian Public Policy 2017 Nov 01;11(3):316-333. [doi: 10.1080/17516234.2017.1396953]
11. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
12. Han Y, Lie RK, Guo R. The internet hospital as a telehealth model in China: Systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
13. Zhou L, Wu Q, Liao B, Wang X, Su N, Yue R, et al. Current situation and development ideas of internet hospitals. Chinese Hospital Management 2019 Nov;39(11):58-60.
14. Chang Z, Chen M. Analysis of medical service mode and the trend of Internet hospital. Chinese Journal of Health Informatics and Management 2016 Dec;13(6):557-560. [doi: 10.3969/j.issn.1672-5166.2016.06.03]
15. Gu Y, Wei D, Cao X, Zhang C, Feng X. Analysis of the reason of the internet hospital occurred in China. Chinese Health Service Management 2017;34(6):401-403.
16. He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, et al. Using an internet-based hospital to address maldistribution of health care resources in rural areas of Guangdong Province, China: Retrospective and descriptive study. JMIR Med Inform 2018 Dec 21;6(4):e51 [FREE Full text] [doi: 10.2196/medinform.9495] [Medline: 30578195]
17. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: Multicenter user profiling study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
18. Li L, Liu G, Xu W, Zhang Y, He M. Effects of internet hospital consultations on psychological burdens and disease knowledge during the early outbreak of COVID-19 in China: Cross-sectional survey study. J Med Internet Res 2020 Aug 04;22(8):e19551 [FREE Full text] [doi: 10.2196/19551] [Medline: 32687061]
19. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: Cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
20. LAI Y, Chen X, Chen S, Huang Y, Huang H, Ung C, et al. PNS345 business model innovation of internet hospitals in China: A systematic review. Value Health 2019 Nov;22:S822. [doi: 10.1016/j.jval.2019.09.2245]

XSL•FO
RenderX

21. Notice on printing and distributing three documents including the internet medical consultation administrative measures (Trial). National Health Commission of the People's Republic of China and National Administration of Traditional Chinese Medicine of the People's Republic of China. 2018 Jul 17. URL: http://www.nhc.gov.cn/yzygj/s3594q/201809/c6c9dab0b00c4902a5e0561bbf0581f1.shtml [accessed 2020-03-10]
22. State Council's opinion about promoting the development of "Internet + Medical Care and Health". General Office of the State Council of the People's Republic of China. 2018 Apr 25. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-03-10]
23. Notice on strengthening information to support the prevention and control work of the pneumonia caused by novel coronavirus infection. National Health Commission of the People's Republic of China. 2020 Feb 03. URL: http://www.gov.cn/zhengce/zhengceku/2020-02/05/content_5474692.htm [accessed 2020-03-10]
24. Notice on promoting internet medical consultation services in epidemic prevention and control. National Health Commission of the People's Republic of China. 2020 Feb 06. URL: http://www.nhc.gov.cn/yzygj/s7653p/202002/ec5e345814e744398c2adef17b657fb8.shtml [accessed 2020-03-10]
25. Guidance of advancing the "Internet +" healthcare insurance service in novel coronavirus pneumonia prevention and control period. National Healthcare Security Administration of the People's Republic of China, National Health Commission of the People's Republic of China. 2020 Mar 02. URL: http://www.nhsa.gov.cn/art/2020/3/2/art_37_2750.html [accessed 2020-03-10]
26. Provisional regulation on the approval of internet drug business. State Food and Drug Administration of the People's Republic of China. 2005 Sep 29. URL: http://www.gov.cn/govweb/fwxx/bw/spypjgj/content_505677.htm [accessed 2020-03-10]
27. Notice to strengthen the administration of internet drug sales. China Food and Drug Administration of the People's Republic of China. 2013 Oct 29. URL: http://www.gov.cn/gzdt/2013-10/29/content_2517631.htm [accessed 2020-03-10]
28. Opinions of National Health and Family Planning Commission on promoting telemedicine service in medical institutions. National Health and Family Planning Commission of the People's Republic of China. 2014 Aug 21. URL: http://www.nhc.gov.cn/yzygj/s3593g/201408/f7cbfe331e78410fb43d9b4c61c4e4bd.shtml [accessed 2020-03-10]
29. Planning of the General Office of the State Council on the national medical and health service system (2015-2020). General Office of the State Council of the People's Republic of China. 2015 Mar 06. URL: http://www.gov.cn/zhengce/content/2015-03/30/content_9560.htm [accessed 2020-03-10]
30. Guidance of the State Council on promoting "Internet +" action. State Council of the People's Republic of China. 2015 Jul 01. URL: http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm [accessed 2020-03-10]
31. Guidance of the General Office of the State Council on promoting tiered healthcare delivery system. General Office of the State Council of the People's Republic of China. 2015 Sep 08. URL: http://www.gov.cn/zhengce/content/2015-09/11/content_10158.htm [accessed 2020-03-10]
32. Guidance of the State Council on promoting and standardizing the application and development of healthcare big data. General Office of the State Council of the People's Republic of China. 2016 Jun 21. URL: http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm [accessed 2020-03-10]
33. Healthy China 2030 plan. The CPC Central Committee and the State Council of the People's Republic of China. 2016 Oct 25. URL: http://www.gov.cn/xinwen/2016-10/25/content_5124174.htm [accessed 2020-03-10]
34. Notice of the State Council on the 13th five-year plan for healthcare. State Council of the People's Republic of China. 2016 Dec 27. URL: http://www.gov.cn/zhengce/content/2017-01/10/content_5158488.htm [accessed 2020-03-10]
35. Notice of National Health and Family Planning Commission on the national population health informatization development in the 13th five-year. National Health and Family Planning Commission of the People's Republic of China. 2017 Jan 24. URL: http://www.nhc.gov.cn/guihuaxxs/s10741/201702/ef9ba6fbe2ef46a49c333de32275074f.shtml [accessed 2020-03-10]
36. Guidance of the General Office of the State Council on promoting the construction and development of medical alliances. General Office of the State Council of the People's Republic of China. 2017 Apr 23. URL: http://www.gov.cn/zhengce/content/2017-04/26/content_5189071.htm [accessed 2020-03-10]
37. Regulatory practices for the registration of physician practices. National Health and Family Planning Commission of the People's Republic of China. 2017 Feb 28. URL: http://www.nhc.gov.cn/fzs/s3576/201808/a359f40edf5e43a79d0174437fdca02f.shtml [accessed 2020-03-10]
38. Action plan for further improvement of medical services (2018-2020). National Health and Family Planning Commission of the People's Republic of China and National Administration of Traditional Chinese Medicine of the People's Republic of China. 2017 Dec 29. URL: http://www.gov.cn/gongbao/content/2018/content_5299607.htm [accessed 2020-03-10]
39. Notice to further develop "Internet + Medical Care and Health" to benefit the people. National Health Commission of the People's Republic of China and National Administration of Traditional Chinese Medicine of the People's Republic of China. 2018 Jul 10. URL: http://www.gov.cn/zhengce/zhengceku/2018-12/31/content_5435186.htm [accessed 2021-03-10]
40. Notice of National Health Commission to carry out pilot work of "Internet + Nursing Service". National Health Commission of the People's Republic of China. 2019 Jan 22. URL: http://www.nhc.gov.cn/yzygj/s7657g/201902/bf0b25379ddb48949e7e21edae2a02da.shtml [accessed 2020-03-10]

XSL•FO
RenderX

41. Law of the People's Republic of China on pharmaceutical administration. National People's Congress of the People's Republic of China. 2019 Aug 26. URL: http://www.gov.cn/xinwen/2019-08/26/content_5424780.htm [accessed 2020-03-10]
42. Guidance of the National Healthcare Security Administration on improving the "Internet +" medical service price and medical insurance payment policy. National Healthcare Security Administration of the People's Republic of China. 2019 Aug 30. URL: http://www.nhsa.gov.cn/art/2019/8/30/art_37_1707.html [accessed 2020-03-10]
43. Notion of multi sectors of action plan on promoting high quality development of health industry (2019-2022). National Development Reform Commission of the People's Republic of China, Ministry of Education of the People's Republic of China, Ministry of Science and Technology of the People's Republic of China, Ministry of Industry and Information Technology of the People's Republic of China, Ministry of Civil Affairs of the People's Republic of China, and Ministry of Finance of the People's Republic of China. 2019 Aug 28. URL: http://www.gov.cn/xinwen/2019-09/30/content_5435160.htm [accessed 2020-03-10]
44. Liu GG, Vortherms SA, Hong X. China's Health Reform Update. Annu Rev Public Health 2017 Mar 20;38:431-448. [doi: 10.1146/annurev-publhealth-031816-044247] [Medline: 28125384]
45. Zhang X, Sun F, Fang P. A cognition survey of grassroots medical staff on the implementation of hierarchical diagnosis and treatment system: a case study of Wuhan. Chinese Rural Health Service Administration 2020 Feb;40(2):82-86.
46. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. Int J Equity Health 2017 Mar 03;16(1):42 [FREE Full text] [doi: 10.1186/s12939-017-0543-9] [Medline: 28253876]
47. Chen K, Han Y, Wang Y, Meng Q. Theoretical framework, implementation path and evaluation system of the tiered diagnosis and medical treatment in China: a literature analysis. Chinese Journal of Public Health 2019 Apr;35(4):497-503. [doi: 10.11847/zgggws1121947]
48. Wang M, He Y, Zhang K, He D, Ynag L, Yu G. Comparison on operation modes of internet hospitals in China. Chinese Health Resources 2020 Mar;23(2):110-113. [doi: 10.13688/j.cnki.chr.2020.20240]
49. A comprehensive analysis of 497 internet hospitals. Vcbeat. 2020 May 07. URL: http://mp.ofweek.com/smarthospital/a045693828696 [accessed 2020-06-07]
50. Zhang X, Zhang X, Yang S, Wang Y. Factors influencing residents' decision to sign with family doctors under the new health care reform in China. Int J Health Plann Manage 2019 Oct;34(4):e1800-e1809. [doi: 10.1002/hpm.2896] [Medline: 31429111]
51. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: A new health care mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
52. Ji L, Liu Z, Yuan Y, Li J, Li X. Analysis on the competition situation of Internet hospitals and countermeasures: the analysis of Wuzhen Internet Hospital. Chinese Journal of Health Informatics and Management 2018 Feb;15(1):105-110. [doi: 10.3969/j.issn.1672-5166.2018.01.021]
53. Johnson MW, Christensen CM, Kagermann H. Reinventing your business model. Harvard Business Review. 2008 Dec. URL: https://hbr.org/2008/12/reinventing-your-business-model [accessed 2021-06-24]
54. China statistical yearbook 2019. National Bureau of Statistics of China. URL: http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm [accessed 2021-06-24]
55. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
56. Zhang W, Ung COL, Lin G, Liu J, Li W, Hu H, et al. Factors contributing to patients' preferences for primary health care institutions in China: A qualitative study. Front Public Health 2020 Aug 18;8:414 [FREE Full text] [doi: 10.3389/fpubh.2020.00414] [Medline: 33014959]
57. Zhang M, Wang Y, Qian Z, Wang D. Analysis on the development model of internet hospitals in China. Health Economics Research 2019 May;36(5):23-26. [doi: 10.14055/j.cnki.33-1056/f.2019.05.007]

Edited by R Kukafka, G Eysenbach; submitted 08.07.20; peer-reviewed by R Guo, X Xie; comments to author 28.08.20; revised version received 22.10.20; accepted 25.05.21; published 20.07.21.

Please cite as:
Lai Y, Chen S, Li M, Ung COL, Hu H
Policy Interventions, Development Trends, and Service Innovations of Internet Hospitals in China: Documentary Analysis and Qualitative Interview Study
J Med Internet Res 2021;23(7):e22330
URL: https://www.jmir.org/2021/7/e22330
doi:10.2196/22330
PMID:34283025

XSL•FO
RenderX

©Yunfeng Lai, Shengqi Chen, Meng Li, Carolina Oi Lam Ung, Hao Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding “Internet Plus Healthcare” in China: Policy Text Analysis

Feng Yang[1], BA, MA, PhD; Huilin Shu[1], BA, MA; Xiaoqian Zhang[2], BA, MA

[1]Sichuan University, Chengdu, China

[2]McGill University, Montreal, QC, Canada

Corresponding Author:
Xiaoqian Zhang, BA, MA
McGill University
3661 Peel St.
Montreal, QC, H3A 1X1
Canada
Phone: 1 5149674660
Email: xiaoqian.zhang@mail.mcgill.ca

Abstract

Background: The combination of the internet and healthcare has excellent benefits and far-reaching positive effects in improving service efficiency and promoting social equity. The role of the “internet plus healthcare” (IPHC) has been recognized, especially during the COVID-19 pandemic. This new healthcare model is also familiar to people and shows a bright prospect.

Objective: This article seeks to accurately understand and fully grasp the characteristics of IPHC policies that can enlighten the formulation of future policies.

Methods: The content analysis method was used to analyze China’s IPHC policies collected from the Beida Fabao database and several official websites.

Results: We found that the development of IPHC policy has gone through 4 stages and is currently entering a phase of rapid development. IPHC policymakers are primarily health administrative departments. In addition, policy instruments are classified into either supply, environment, or demand, and policy themes can be summarized into 4 categories: facilities, technology, service, and management.

Conclusions: China’s IPHC policy has good prospects from the perspective of development trends. The health administrative departments mainly lead the development of China’s IPHC policy. It is suggested that these departments involve other stakeholders (ie, medical workers, medical industries, and technology sectors) in formulating policies. Policies prefer to use supply-based and environment-based policy instruments. The policy themes emphasize improving infrastructure construction and high-quality diagnostic and treatment services, strengthening the supporting role of information technology, and ensuring all stakeholders understand their responsibilities.

(J Med Internet Res 2021;23(7):e23779) doi:10.2196/23779

KEYWORDS

internet plus health care; China; policy analysis; COVID-19; epidemic

Introduction

The internet has become a driving force for different fields, including healthcare, education, and entertainment. To maximize the potential of the internet, the Chinese government has created “internet plus” designs to transform, modernize, and equip traditional industries. Likewise, the healthcare industry has been combined with the internet to create “internet plus healthcare” (IPHC).

IPHC is a novel application of the internet in the healthcare industry that includes health education, medical information queries, electronic health records or electronic medical records (EHR/EMRs), disease risk assessments, online disease consultations, electronic prescriptions, remote consultations, and various remote forms of health and medical services such as treatment and rehabilitation. Chinese hospitals use the internet and mobile technologies to alleviate the challenges patients encounter when obtaining hospital services [1]. Since the outbreak of COVID-19, IPHC has played a role in “high

XSL•FO
RenderX

efficiency and low risk" health care delivery [2] by enabling a more efficient pandemic response and launching a variety of valuable services such as addressing pandemic queries in real-time, facilitating online consultations, and providing home isolation guidelines. IPHC also minimizes the risk of cross-infection via in-person consultations. As a result, the IPHC model underwent significant development in 2020. For example, Alibaba Health launched a free clinic service on the internet, and Jingdong Health launched the "Preventing and Blocking COVID-19 Pneumonia" platform.

The development of IPHC relies on effective national policies. Policies reflect conceptualization, subjectivity, and practice in specific fields. Policy analysis enables an understanding of previous policies and offers retrospective and prospective insights for future policy development and implementation [3]. It offers a robust path to understanding how and why governments enact certain policies, including their values, interests, and the political contexts [4]. As a solid reflection of public affairs and the essence of policy content, policy text has become the primary starting point for policy analysis. Converting the policy text into several elements allows for a better examination of themes contained in the policy. Health policy analysis can potentially resolve protracted policy disputes and strengthen a sustainable health system [5].

Existing research on health policy in China either focuses on reformation [6-8] or tries to assess how specific diseases affect China's health policy [9,10]. Previous studies evaluated the combination of the internet and healthcare in China by studying internet hospitals. For example, Xie et al [11] provided an overview of the internet hospitals in China; Han et al [12] analyzed the construction and content of internet hospitals in China. In addition, the number of Chinese citizens who use the internet to seek health services is increasing; therefore, there is an urgent need to update and refine the current IPHC policies [13]. Several health policy analysis frameworks have been proposed, including the policy triangle framework [14] and the network frameworks [15]. Considering "researchers need to use existing frameworks and theories of the public policy process more extensively [3]," the analysis of IPHC policies will combine the policy triangle framework and the network frameworks with practical requirements. This paper will address the following research questions, which contain elements proposed by previous research and can be collected from policy texts:

1. When are these policies issued?
2. Who issues these policies?
3. What policy instruments are included in these policies?
4. What are these policy themes?

By answering these questions, this paper aims to track the policy trajectory of IPHC in China, present the instruments guaranteed for policy implementation, and understand the key themes of IPHC.

Methods

Research Materials

Policy text was collected in 3 steps from several sources published on May 10, 2020. First, we used the Beida Fabao database, one of China's most professional legal databases, characterized by rich content, detailed classification, and timely data. The retrieval strategy was to obtain relevant policy text with keywords, including "internet medical," "network medical," "internet health," and "network health" through a full-text search. Second, we retrieved new keywords in the obtained policy text, such as "electronic health records," "telemedicine," and "internet hospital," as clearly stated in the "Opinions of the General Office of the State Council on Promoting the Development of 'Internet plus Healthcare.'" Finally, we supplemented the retrieval strategy with policy text published on the official websites of affiliated agencies such as the National Health Commission of the People's Republic of China (NHC) and the National Administration of Traditional Chinese Medicine (SATCM). Given that the content of local IPHC policies was mainly based on national policies, and the focus of local policy themes was consistent with national policies, IPHC policies issued by the local government were eliminated from the analysis. Likewise, duplicate policy and invalid policy documents were removed, resulting in 90 policy texts.

Research Method

The content analysis method was used to convert a large amount of text into a small number of categories to help researchers discover the policy development patterns and trends [16]. The content analysis was designed to examine the text and explain "what they mean to people, what they enable or prevent, and what the information conveyed by them does" [17]. The content analysis method identified IPHC policy codes and classified the textual information per the following criteria:

The Evolution of Policies

The issuance period shows the development trend of the IPHC policies. We coded the date of policy issuance into annual units.

Policymakers

The coding for issuing markers to government agencies enables us to distinguish which government agencies are mainly involved in policymaking and which government agencies cooperate more in the policymaking process. We coded the policymakers by collecting their names as written in the policy issuing agencies sections of the policy texts. Then we standardized the names of policymakers according to the latest 2018 State Council Institutional Reform Plan.

Policy Instruments

Based on the classification of policy instruments by Rothwell and Zegveld [18], policy instruments in IPHC policies were classified into supply-based (eg, "technology and infrastructure" and "education and training"), environment-based (eg, "goal planning" and "legal supervision"), and demand-based policy instruments (eg, "medical insurance system" and "online services"). For example, the "Notice of the State Administration of Traditional Chinese Medicine on Printing and Distributing

XSL•FO
RenderX

the Twelfth Five-Year Plan for the Informatization of Traditional Chinese Medicine" can be coded as "integrated medical," "information support," "internet-based Chinese medicine," and "education and training." According to this classification, 2 coders coded the 90 policy texts.

Policy Themes

Policy themes can reflect the significant concerns derived from policy documents. We retrieved keyword segmentation results using the Jieba for Python and analyzed high-frequency keywords. We formulated the co-occurrence matrix of high-frequency keywords to examine the main clusters of IPHC themes from the textual information. We used Gephi to visualize the policy themes.

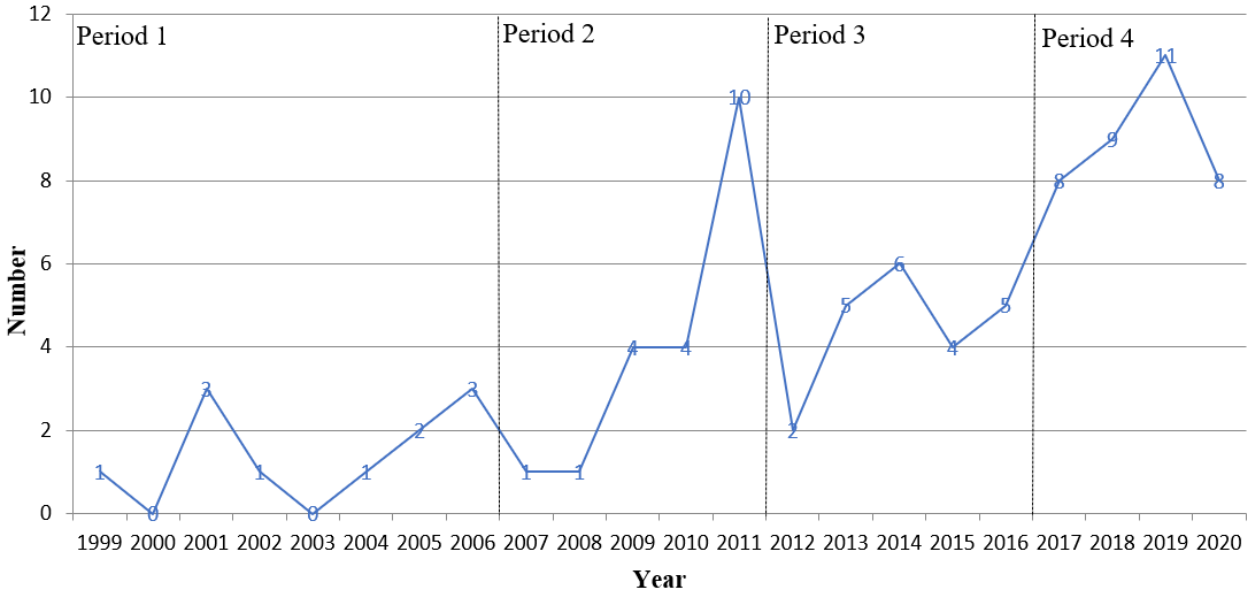
The consistency of the coding results by 2 coders was tested to ensure the reliability and objectivity of the content analysis of IPHC policy documents. The formula which was used to determine reliability was consistency coefficient =  (where M is the number of consistent codes derived from the 2 coders, and N1 and N2 are the coder's coding numbers). The calculated result showed that the consistency coefficient of coding was within an acceptable range (0.854).

Results

The Evolution of Policies

As an essential part of China's national economic plan, The Five-Year Plan makes long-term plans for China's major construction projects. It specifies the goals and directions for national economic development. According to the 5-year plans for the development of health services issued by the State Council (SC; "Notice of the State Council on Approving and Transmitting the Outline of the 'Eleventh Five-Year Plan' for the Development of Health Services," on May 21, 2007; "Notice of the State Council on Printing and Distributing the 'Twelfth Five-Year Plan' for the Development of Health Services," on October 8, 2012; and "Notice of the State Council on Printing and Distributing the 'Thirteenth Five-Year' Health and Wellness Plan," on December 27, 2016), the development of IPHC policy can be divided into four stages (Figure 1), namely, "The Budding Period" (1999-2006), "The Initial Development Period" (2007-2011), "The Reform and Development Period" (2012-2016), and "The Rapid Development Period" (2017-present).

Figure 1. Time distribution of "internet plus healthcare" policy.



Policymakers

The results of IPHC policymakers are shown in Table 1. Results showed that 24 departments issued IPHC policies, including NHC, SATCM, National Medical Products Administration (NMPA), and National Development and Reform Commission. The NHC is the leading organization and has published the most policies (49 policies). This is because the main responsibility of the NHC is to organize the formulation of national health policies, laws, regulations, plans for health development, and departmental rules and standards and organize their implementation. Of the 90 IPHC policy texts, 73 (81.11%) policies were promulgated independently by 1 department. There are 17 joint promulgated policies, and the NHC participates in each promulgated joint policy. The 15 departments represented by The State Council Information Office of the People's Republic of China, Ministry of Education of the People's Republic of China (MOE), and Ministry of Finance of the People's Republic of China (MOF) have promulgated one policy each, all of which are issued jointly.

XSL•FO
RenderX

Table 1. Statistics of each IPHC policy-issuing department.

Policymakers	Number of policies	Joint number of policies	Frequency of joint policies (%)
National Health Commission of the People's Republic of China	49	17	34.69
National Administration of Traditional Chinese Medicine	27	9	40.91
National Medical Products Administration	19	2	7.14
National Development and Reform Commission	4	4	100
State Council	4	0	0.00
Ministry of Industry and Information Technology of the People's Republic of China	3	3	100
National Healthcare Security Administration	3	2	66.67
Ministry of Science and Technology of the People's Republic of China	3	3	100
Ministry of Civil Affairs of the People's Republic of China	2	2	100
The State Council Information Office of the People's Republic of China	1	1	100
Ministry of Education of the People's Republic of China	1	1	100
Ministry of Finance of the People's Republic of China	1	1	100
Ministry of Human Resources and Social Security of the People's Republic of China	1	1	100
Ministry of Natural Resources of the People's Republic of China	1	1	100
Ministry of Ecology and Environment of the People's Republic of China	1	1	100
Ministry of Housing and Urban-Rural Development of the People's Republic of China	1	1	100
Ministry of Commerce of the People's Republic of China	1	1	100
Ministry of Culture and Tourism of the People's Republic of China	1	1	100
The People's Bank of China	1	1	100
State Taxation Administration of the People's Republic of China	1	1	100
State Administration for Market Regulation	1	1	100
General Administration of Sport of China	1	1	100
China Banking and Insurance Regulatory Commission	1	1	100
China Securities Regulatory Commission	1	1	100

Policy Instruments

Policy instruments reflect the intent of policy formulation and determine its effectiveness. They are selected based on economic, political, and social contexts [19], but they are rarely mentioned in public health literature [20]. This study refers to the classification of policy instruments by Rothwell and Zegveld [18], which includes supply-based policy instruments (public enterprise, scientific and technical, education, and information), demand-based policy instruments (procurement, public service, commercial, and overseas agents), and environment-based policy instruments (political, legal and regulation, taxation, and financial). The constructed analysis framework of the IPHC policy instruments is shown in Textbox 1. Supply-based instruments ensure the government directly supports various stakeholders with technology, infrastructure, talent, and resources, thereby promoting IPHC development. Environment-based policy instruments create a suitable environment for developing the IPHC industry by improving laws and regulations and supervisory protocols with clear inspection systems. Finally, demand-based policy instruments explore the demand for IPHC services and promote its development through the medical insurance system and online services.

Results of statistics on the number of policies of various types of IPHC policy instruments are shown in Table 2. Overall, the different types of policy instruments are involved. However, supply-based and environment-based policy instruments are used more, while demand-based policy instruments are used less. Among the supply-based policy instruments, the most used is the "information support," accounting for 47.78% of the 90 policies, followed by the "management" and "legal supervision," which account for 38.89% and 35.56% of environment-based

XSL•FO
RenderX

policy instruments, respectively. The least used policy instrument is the "performance assessment" (8.89%).

Textbox 1. Analysis framework of IPHC policy instruments.

Supply-based

- Technology and infrastructure: Provide the necessary infrastructure and technology to develop IPHC.
- Education and training: Carry out various education and training activities for practitioners involved in IPHC, provide learning resources, and strengthen the development of relevant talents.
- Information support: Build related databases and knowledge bases and make full use of information technologies to provide information exchange and information services to develop IPHC.
- Resource allocation: Medical resources, financial subsidies, and other resources are allocated according to the characteristics and needs of the development of IPHC.

Environment-based

- Goal planning: The development direction of IPHC is defined by formulating macro targets and overall planning.
- Legal supervision: IPHC stakeholders are regulated by establishing regulations, laws, and industry standards.
- Management: The quality of services or products provided by medical and health institutions and related enterprises is managed.
- Policy publicity: IPHC policy is publicized actively to expand the beneficiary groups of the policy and generate a positive public opinion.
- Entities collaboration: The relevant entities of IPHC are encouraged to collaborate and actively promote IPHC development.
- Performance assessment: A scientific performance assessment system is given to subjects offering IPHC information and services.

Demand-based

- Medical insurance system: Users' rights and interests in IPHC activities are fully protected through continuous improvement and optimization of the medical insurance system.
- Online services: Traditional offline consultation and sales services are transformed into convenient and reliable online services through internet technology platforms.
- Integrated medical: Users are offered a more convenient medical experience through the interconnection of medical and health information resources and the full integration of high-quality medical resources.
- Internet-based Chinese medicine: The integrated development of Chinese medicine health service and the internet should be promoted.

Table 2. Statistics of IPHC policy instruments.

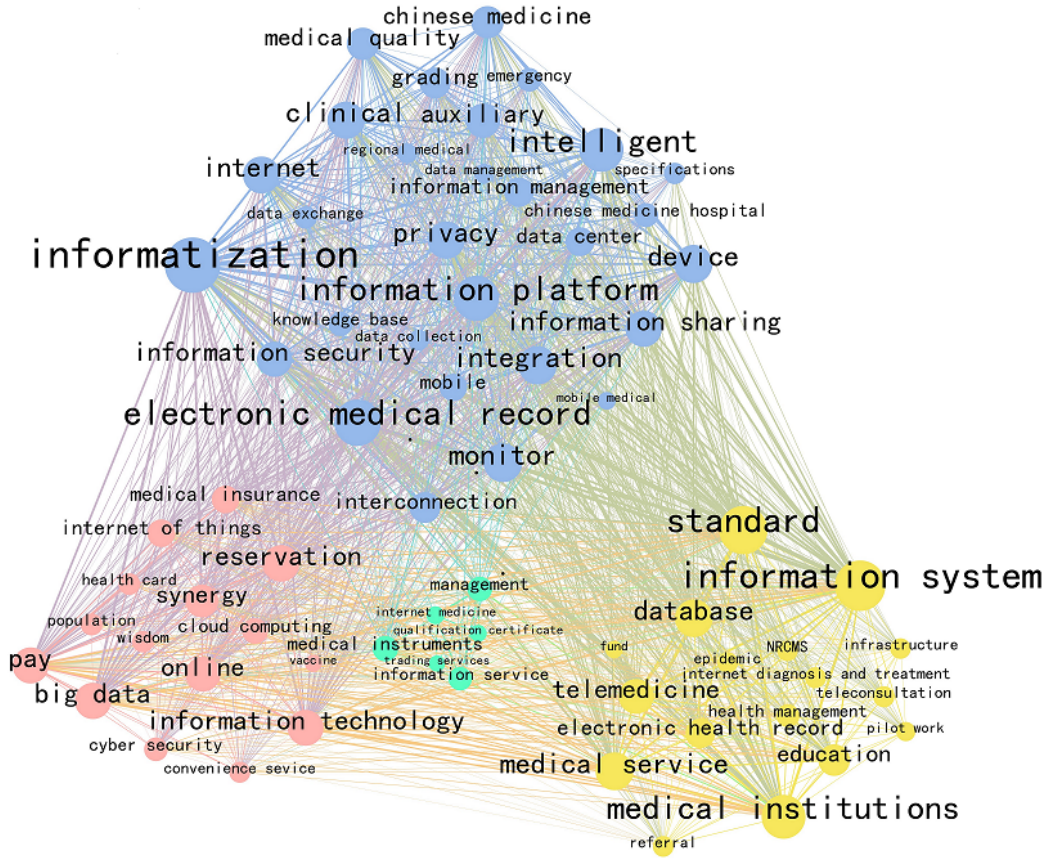
Policy instruments	Number	Frequency (%)
Supply-based		
Technology and infrastructure	27	30.00
Education and training	18	20.00
Information support	3	47.78
Resource allocation	11	12.22
Environment-based		
Goal planning	15	16.67
Legal supervision	32	35.56
Management	35	38.89
Policy publicity	11	12.22
Entities collaboration	18	20.00
Performance assessment	8	8.89
Demand-based		
Medical insurance system	12	13.33
Online services	16	17.78
Integrated medical	21	23.33
Internet-based Chinese medicine	9	10.00

XSL•FO
RenderX

Policy Themes

A total of 4586 keywords were extracted using Jieba for Python. The T value was approximately 37 based on the high-low frequency words boundary fraction formula proposed by Donohue [21], where T is the lowest frequency among the high-frequency keywords, and I_1 is the number of keywords with high frequency. There were 1107 keywords with a frequency greater than 37, and a visualization map with high-frequency words was generated using Gephi. Gephi is a data visualization software that can cluster points of the same attribute (Figure 2). As a result, the themes of IPHC policy can be summarized into 4 categories: facilities (blue), technology (pink), services (yellow), and management (green).

Figure 2. The visualization map of high-frequency keywords.



Discussion

Principal Findings

The study found that IPHC policy can be divided into 4 phases: “The Budding Period” (1999-2006), “The Initial Development Period” (2007-2011), “The Reform and Development Period” (2012-2016), and “The Rapid Development Period” (2017-present). In addition, IPHC policymakers are led by the health administrative department. As the core department, the NHC has formulated the most policies and cooperates extensively with MOE, MOF, and other departments. Furthermore, China’s IPHC policy instruments cover 3 types: supply, environment, and demand, with a preference for the supply-based and environment-based policy instruments. Finally, IPHC policy themes can be summarized into 4 categories: facilities, technology, services, and management, emphasizing the improvement of infrastructure construction, strengthening the supporting role of information technology, improving high-quality diagnosis and treatment services, and ensuring all stakeholders understand their responsibilities.

The Prospect of Policy Development

An early prototype of China’s IPHC appeared in the 1990s when doctors in China began to communicate with experts in other countries through emails about clinically incurable diseases. After that, with the increasing use of computers for remote medical consultations in various places, NHC issued the “Notice on Strengthening the Management of Telemedicine Consultation” in 1999 to regulate medical order and medical behavior and enable the development of healthy and orderly telemedicine consultation work. However, general interest in information technology and internet use among the public was limited at that time. Thus, the number of related policies released during that period was relatively small. At this stage, the largest number of policies was promulgated in 2001. The 3 policies issued in 2001 all focused on the “Internet Drug Information Service” and the regulation of drugs obtained online.

With the innovation of information technology and the widespread popularity of information networks, SATCM issued the “Outline of the Eleventh Five-Year Plan for Chinese Medicine Informatization” in 2007, emphasizing the

XSL•FO
RenderX

construction of various information systems in healthcare. At the end of 2010, NHC issued the "Electronic Medical Record System Functional Specification (Trial)," which regulated the management of EHR/EMRs in medical institutions, promoted the construction of hospital informatization, and profoundly affected subsequent policy development. For example, the number of policies issued in 2011 peaked at this stage, and most of them emphasized the construction of hospital informatization based on EHR/EMRs. Overall, the integration of healthcare and the internet ushered in a preliminary development stage. As a result, the relevant policies issued were further improved, with policies covering a more comprehensive range of topics. In addition, the basic information database was completed, the medical management information system was continuously improved, and talent training programs in the medical field were initiated.

In 2011, many medical institutions and doctors began to use Weibo (a popular social media app in China) to interact with patients. However, with the ensuing privacy exposure and medical disputes, government departments began to regulate and manage social media–based medical services, and IPHC entered a period of deepening reform. Overall, more policies were issued at this stage compared to the previous stage, and the number of policies fluctuated less frequently. Various targeted measures were introduced to IPHC to resolve issues regarding an imperfect management system and operation mechanism, the lack of information infrastructure development, low levels of information technology application, and the insufficient sharing of medical information resources. For example, specific policies have been issued for telemedicine services, information system development, information technology training, and internet pharmaceutical sales. In addition, the Communist Party of China (CPC) and the SC jointly issued "Healthy China 2030" in 2016, which outlines the blueprint and action plan for the construction of a healthy China over the next 15 years, proposing to regulate and promote IPHC services, and innovate the IPHC service model. The introduction of this outline was the first time that the country clearly stated its attitude toward IPHC and elevated it to the level of a national strategy.

With the reform and development of the previous stage, China's healthcare has made significant progress, and people's health has continued to improve. Therefore, the CPC and SC have attached great importance to the development of health and wellness and placed it in a prominent position for economic and social development. In the 19th National Congress of the CPC report on October 18, 2017, the development strategy of "healthy China" was put forward, and China's IPHC policy entered a period of rapid development. In 2018, in response to the problems encountered in the vigorous development of IPHC, the General Office of the SC issued "Opinions on Promoting the Development of 'internet plus healthcare,'" proposing a series of policies to promote the integration of the internet and medical health and urge all departments to issue supporting policy measures promptly. Furthermore, as a result of COVID-19, the NHC successively issued several departmental regulatory documents before May 2020 to leverage the complete advantages of internet healthcare, which has provided robust support for epidemic prevention and control. As such, the pandemic has promoted the development of IPHC to a certain extent, and the number of related policies is expected to increase significantly in the near future.

The Multiple Voices in Policymaking

The "internet plus" industry, including IPHC, is policy-oriented in China. Compared to other industries, the IPHC field is more professional and requires complete, detailed, and operable policy documents to regulate industry behavior and guide industry development. Overall, the current IPHC policy issuance is mainly led by the health administrative department. For example, the NMPA is responsible for internet-based drug-related policies, and the NHC coordinates industry development and formulates industry standards. In addition to continuing its primary role in future policy formulation, it strengthens the coordination with nonhealthy departments to leverage the advantages of each department. It is suggested that these departments should be involved in formulating policies for various social services. They should listen to the voices of different interest groups in the IPHC field, coordinate conflicts of interest, and enhance the enforceability of policies. For example, medical personnel and the medical industry offer a unique perspective in formulating health policies because of their knowledge, technology, and position [3]. In addition, the health administration should strengthen communication and cooperation with internet companies, telecommunications industries, communities, and the general public.

The Coordination of Various Policy Instruments

"Technology and infrastructure" account for a relatively high proportion (30%) of supply-based policies; however, the importance of such policies on emerging technologies needs to be improved. Nambisan [22] found healthcare organizations that provide patients with online health communities need to pay more attention to developing tools that will make internet searches more effective. Considering the role of cloud computing, big data, artificial intelligence, 5G, and other emerging technologies in promoting the positive development of IPHC, future policies need to closely follow technology development trends and provide forward-looking and innovative guidance for the development of IPHC.

"Education and training" accounts for 20% of all policy instruments and can be divided into 2 categories: continuing education for medical practitioners and providing a training program for the new personnel joining the IPHC industry. Compared to the former, the latter has had a late start, and its policy content is primarily macro; thus, it is necessary to provide detailed training programs for new talent. The number of policies using "information support" covers the largest proportion (47.78%) of supply-based policy instruments. They all emphasize improving the level and ability of medical information services and actively creating a development culture for the growth of IPHC through the construction of information platforms and other informatization means. The "resource allocation" policy instruments are used less frequently, accounting for only 12.22%; hence, future policies should consider further strengthening financial investments.

XSL•FO
RenderX

Among the environment-based policy instruments, "management" is used the most (39.89%). It is mainly focused on EHR/EMRs, telemedicine, and internet drug information services. The least used is "performance assessment" (8.89%) and is reflected in telemedicine pilots, informatization of traditional Chinese medicine and vaccines, and electronic registration after 2015. This type of policy instrument needs to be improved further to provide a scientific performance system to assess the management of relevant practitioners. "Entities collaboration" accounts for 20% of environment-based policy instruments, mainly emphasizing the coordination of supervision among policy issuers and the cooperation of medical institutions; however, participation and collaboration among other social services are relatively neglected. The policies using "policy publicity" account for 12.22% of environment-based policies. Most of them were published after the "Guiding Opinions of the General Office of the State Council on Promoting and Regulating the Application of Health and Medical Big Data" was issued by the General Office of SC in 2016. This policy instrument needs to be used continuously to create a suitable environment for engaging public opinion in developing IPHC.

"Integrated medical" (23.33%) accounts for a large proportion of demand-based policy instruments, and it was developed after 2014. The second is "online services" (17.78%), which emphasized the development of "internet plus diagnosis and treatment" and "internet plus pharmaceuticals," the interconnection of medical resources, and the optimization of the medical experience. Only 13.33% of the demand-based policy instruments cover the "medical insurance system," which is essential to protecting people's health. In addition to the previous medical insurance payment guidance on telemedicine, other policies have begun to use the "medical insurance system" following the "Notice of the General Office of the National Health and Family Planning Commission on the Comprehensive Promotion of the Construction of the National New Rural Cooperative Medical Information Platform" issued in 2015. This policy instrument needs to be improved in terms of pricing instructions, pricing standards, and service areas to effectively solve user needs, dispel user concerns, and promote the development of IPHC. "Internet-based Chinese medicine" accounts for 10% of all demand-based policies, emphasizing the development of the internet plus Chinese medicine health services; however, specific and feasible solutions are still under discussion.

The Four Categories of the Policy Themes

"Facilities" focuses on the infrastructure (eg, the construction and improvement of network equipment, network environment, databases, information systems, and mobile devices) required during the development of IPHC, utilizing core keywords such as "informatization" and "electronic medical records." "Informatization is a relatively abstract concept related to the combination of medical treatment and information technology, using information technologies such as the internet to improve medical efficiency and service levels. It is referred to in the policy as "implement a national health security information project, expand and improve existing facilities, fully build a shared population health information platform, and strengthen data collection, integrated sharing, and business collaboration of application information systems such as public health, medical services, and drug supply" [23]. For example, relying on China Unicom's 5G plus a medical cloud platform, experts in Beijing, Shanghai, Guangzhou, and other places remotely consulted with critically ill patients in Leishenshan Hospital in Wuhan in February 2020, improving medical efficiency. "Electronic medical record" and "informatization" function in tandem. Existing policies emphasize constructing a basic database with EHR/EMRs as the core of healthcare informatization. They also recommend enhancing the unification and standardization of EHR/EMRs standards to achieve big data management of health information for the entire life cycle of the population. However, EHR/EMRs have not been widely popularized in Chinese hospitals. Traditional paper medical records are still the standard, and the transmission and sharing of EHR/EMRs between hospitals are even more limited. As the foundation of IPHC, the optimization of "facilities" is of vital importance, and China still needs to invest additional efforts to ensure the development of IPHC by improving the network, database, and mobile devices.

The evolution of technology can bring substantial benefits to healthcare [24]; even "policymakers worldwide view information technology as a means of making healthcare systems safer, more affordable, and more accessible" [25]. The "technology" theme of China's IPHC also reiterates the foundational importance of information technology, emphasizing the development of cloud computing, the internet of things, and other information technologies to provide technical support for the development of IPHC. Representative high-frequency keywords in this category include "payment," "online," and "synergy." Among them, "payment" refers to the development of mobile payment technologies (eg, network payment or online payment) to reduce patients' queuing time and optimize hospital service processes. Especially in recent years, mobile payment terminals have become standard equipment in various hospitals, and patients also tend to use mobile apps such as Alipay and WeChat to pay their fees. "Online" emphasizes the conversion of traditional medical treatment processes to online platforms, including online queries, online appointments, online viewing of medical results, online payment, and online drug purchases. At present, many hospitals have initiated their official portals, WeChat public accounts, and other network portals. By visiting these portals, patients can complete the entire process of scheduling appointments, medical treatments, and medication purchases. Even under the influence of this trend, professional internet medical platforms such as Ping An Good Doctor (PAGD), Ding Xiang, Chunyu, and Good Doctor have begun to appear. For example, Ping An Health Cloud Company launched PAGD (an online health consultation and health management app) in 2015, and the number of user registrations exceeded 315 million by the end of 2019 [26]. Unlike traditional hospitals, PAGD provides users with one-stop medical services, including 24/7 online consultations, referrals, registrations, online drug purchases, and 1-hour drug deliveries. "Synergy" is a collective term for "regional synergy," "online and offline synergy," "business synergy," etc. It is believed that the process of developing IPHC should (1) improve coordination between different regions to promote access to high-quality medical resources in remote areas; (2) optimize the medical service

XSL•FO
RenderX

process by coordinating the traditional medical model and online medical model; and (3) coordinate various business-related information, promote the integration of information resources (eg, medical service prices, drug information, and medical insurance payments) and promote the joint reform of medical industry, pharmaceutical industry, and medical insurance system.

The "service" category includes keywords such as "information system," "medical institution," and "medical service," emphasizing the optimization of medical services and the improvement of service levels. Generally speaking, companies can accurately identify consumer needs and quickly integrate resources through the internet to improve services. For hospitals, the use of the internet can provide people with safe, effective, convenient, and inexpensive essential medical services, significantly improve patients' experience with doctor visits, and improve the quality of medical services. As a result, each medical institution should strengthen the use of information technology, and government departments should improve relevant information and technology standards and strengthen medical information education. For example, several policies promulgated during the COVID-19 epidemic have recognized the role of IPHC in responding to sudden infectious diseases, requiring government departments and hospitals at all levels to leverage the beneficial experience of internet-based diagnosis and treatment, the construction of internet hospitals, and the use of telemedicine services during the pandemic. This further promotes the integrated development of internet technology and medical services and the normalization of telemedicine services.

In developing IPHC, problems such as user privacy, qualifications of practitioners, and responsibilities and obligations of both doctors and patients have become prominent. Improper management of these concerns will risk patients' health. At the same time, it will stigmatize IPHC and damage the development of the industry. For example, Bansal et al [27] found that users' trust in medical websites and concerns about privacy affect the extent to which they provide personal health information on the internet, which affects the success of IPHC Therefore, the "management" of patient privacy concerns is necessary. It focuses on the supervision of all stages in the development process of IPHC and specifies the responsibilities of all parties. For example, health administrative departments at all levels should strengthen the supervision and management of internet diagnosis and treatment services, publicize the list of medical institutions that conduct internet diagnosis and treatment activities, and facilitate the supervision of these institutions by the general public.

Furthermore, medical institutions should strengthen the use and management of internet diagnosis and treatment activities, ensure that they are traceable throughout the process, and open data interfaces with regulatory authorities. In addition, physicians who carry out internet diagnosis and treatment activities shall obtain corresponding qualifications according to the law. Currently, government departments at all levels in China have successfully issued specific policies to regulate the development of this industry. For instance, the "Opinions on Promoting the Development of 'Internet plus Healthcare'" issued in 2018 by the SC required the NHC, Cyberspace Administration of China, Ministry of Industry and Information Technology of the People's Republic of China, and other departments to issue management measures to regulate internet diagnosis and treatment, to clarify the standards for supervision, and to ensure the quality and safety of medical and health services. Likewise, the Sichuan Province, Zhejiang Province, and other places have also established provincial level IPHC supervision platforms. These platforms supervise online medical institutions, prescriptions, diagnosis and treatment content, service quality, and the qualifications of doctors, nurses, and pharmacists.

Conclusions

IPHC policy clarifies the development goals of the medical industry and the main contents of industry development. It also promotes the modernization of medical service methods. This article used content analysis to analyze policymakers, policy instruments, and the policy themes of IPHC policy in China, aiming to determine the characteristics of current policies, examine the inherent laws of policies, inspire future policy formulation, and guide the policies of the IPHC industry better.

This article has some limitations. Considering that the local policies were issued under the guidance of the central policy, we only selected the policies at the central level for analysis. However, the study of local policies can effectively show the spread, speed, and influence of policies and identify regional characteristics for future research. In addition, although this article analyzes several aspects of IPHC policy, it is limited by time and effort and cannot further explain the interconnection between these aspects. For example, which policymaker prefers to use which policy instrument, and do different policy themes use other policy instruments? Additional in-depth research is required to answer these questions.

Acknowledgments

This research was based on work funded and supported by the Sichuan University (grant SKSYL201807).

Conflicts of Interest

None declared.

References

1. Tu J, Wang C, Wu S. Using technological innovation to improve health care utilization in China's hospitals: the emerging 'online' health service delivery. Journal of Asian Public Policy 2017 Nov 14;11(3):316-333. [doi: 10.1080/17516234.2017.1396953]

XSL•FO
RenderX

2. Zhang M. Ali Health Aims Higher with Rebranding and New Strategies. 2020. URL: https://equalocean.com/analysis/2020091114753 [accessed 2021-01-14]
3. Walt G, Shiffman J, Schneider H, Murray SF, Brugha R, Gilson L. 'Doing' health policy analysis: methodological and conceptual reflections and challenges. Health Policy Plan 2008 Sep;23(5):308-317 [FREE Full text] [doi: 10.1093/heapol/czn024] [Medline: 18701552]
4. Browne J, Coffey B, Cook K, Meiklejohn S, Palermo C. A guide to policy analysis as a research method. Health Promot Int 2019 Oct 01;34(5):1032-1044. [doi: 10.1093/heapro/day052] [Medline: 30101276]
5. Koon AD, Hawkins B, Mayhew SH. Framing and the health policy process: a scoping review. Health Policy Plan 2016 Jul;31(6):801-816 [FREE Full text] [doi: 10.1093/heapol/czv128] [Medline: 26873903]
6. Qian J, Blomqvist AG. Health policy reform in China: A comparative perspective. Kiriat Hatikshoret-Neve Ilan Suite 226-Harei Yehuda, 90850, ISRAEL: World Scientific Publishing Company; 2014.
7. Ramesh M, Wu X. Health policy reform in China: lessons from Asia. Soc Sci Med 2009 Jun;68(12):2256-2262. [doi: 10.1016/j.socscimed.2009.03.038] [Medline: 19419809]
8. Wang H, Gusmano K, Cao Q. An evaluation of the policy on community health organizations in China: will the priority of new healthcare reform in China be a success? Health Policy 2011;99(1):37-43. [doi: 10.1016/j.healthpol.2010.07.003] [Medline: 20685005]
9. Xu J, Wang J, Wimo A, Fratiglioni L, Qiu C. The economic burden of dementia in China, 1990-2030: implications for health policy. Bull World Health Organ 2017 Jan 01;95(1):18-26. [doi: 10.2471/BLT.15.167726] [Medline: 28053361]
10. Zhang L, Wilson DP. Trends in notifiable infectious diseases in China: implications for surveillance and population health policy. PLoS One 2012;7(2):e31076 [FREE Full text] [doi: 10.1371/journal.pone.0031076] [Medline: 22359565]
11. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
12. Han Y, Lie RK, Guo R. The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
13. Ng F. New model of "Internet plus Healthcare": case study of online health services in China. Israel Journal of Health Policy Research(1). 2019. URL: https://tinyurl.com/ywju5c32 [accessed 2021-01-14]
14. Walt G, Gilson L. Reforming the health sector in developing countries: the central role of policy analysis. Health Policy Plan 1994 Dec;9(4):353-370. [doi: 10.1093/heapol/9.4.353] [Medline: 10139469]
15. Schneider H, Gilson L, Ogden J, Lush L, Walt G. Health systems and the implementation of disease programmes: case studies from South Africa. Global public health 2006;1(1):64. [doi: 10.1080/17441690500361083] [Medline: 19153894]
16. Erlingsson C, Brysiewicz P. A hands-on guide to doing content analysis. Afr J Emerg Med 2017 Sep;7(3):93-99 [FREE Full text] [doi: 10.1016/j.afjem.2017.08.001] [Medline: 30456117]
17. Krippendorff K. Content analysis: An introduction to its methodology (4th ed.). Thousand Oaks, CA: Sage Publications, Inc; 2018.
18. Rothwell R, Zegveld W. Industrial Innovation and Public Policy: Preparing for the 1980s and 1990s. London: Frances Printer; 1981.
19. Mohammad A. Sustainability Assessment Context of Resource and Environmental Policy. Cambridge, Massachusetts: Academic Press; 2012.
20. Shroff MR, Jones SJ, Frongillo EA, Howlett M. Policy instruments used by States seeking to improve school food environments. Am J Public Health 2012 Feb;102(2):222-229. [doi: 10.2105/AJPH.2011.300338] [Medline: 22390436]
21. Donohue JC. Understanding scientific Literature: A Bibliographic Approach. Cambridge: The MIT Press; 1973.
22. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
23. Guiding Opinions of the General Office of the State Council on Promoting and Regulating the Development of Health and Medical Big Data Applications. The State Council. 2016. URL: http://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm [accessed 2021-01-17]
24. Rodrigues JJPC, De Rezende Segundo DB, Junqueira HA, Sabino MH, Prince RM, Muhtadi JAI, et al. Enabling Technologies for the Internet of Health Things. IEEE ACCESS 2018;6:13129-13141. [doi: 10.1109/access.2017.2789329]
25. Agarwal R, Gao G, DesRoches C, Jha AK. Research Commentary-The Digital Transformation of Healthcare: Current Status and the Road Ahead. Information Systems Research 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
26. Ping An Good Doctor. Ping An Good Doctor releases 2019 sustainability report, platform visits reached 1.11 billion during the epidemic. 2020. URL: https://www.jk.cn/aboutUs/news/114 [accessed 2021-01-12]
27. Bansal G, Zahedi FM, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision Support Systems 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]

Abbreviations

CPC: The Communist Party of China

XSL•FO
RenderX

EHR/EMR: electronic health record or electronic medical record
IPHC: internet plus healthcare
MOE: Ministry of Education of the People's Republic of China
MOF: Ministry of Finance of the People's Republic of China
NHC: The National Health Commission of the People's Republic of China
NMPA: National Medical Products Administration
PAGD: Ping An Good Doctor
SATCM: The National Administration of Traditional Chinese Medicine
SC: State Council

Edited by R Kukafka, G Eysenbach; submitted 23.08.20; peer-reviewed by A El Rifay, I Pramukti, W Zhang, S Tang; comments to author 03.12.20; revised version received 17.01.21; accepted 30.06.21; published 26.07.21.

Please cite as:
Yang F, Shu H, Zhang X
Understanding "Internet Plus Healthcare" in China: Policy Text Analysis
J Med Internet Res 2021;23(7):e23779
URL: https://www.jmir.org/2021/7/e23779
doi:10.2196/23779
PMID:34309581

©Feng Yang, Huilin Shu, Xiaoqian Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Consultation Pricing of the Online Health Care Service in China: Hierarchical Linear Regression Approach

Ya-Ling Chiu[1,2,3*], PhD; Jying-Nan Wang[2,3,4*], PhD; Haiyan Yu[5*], PhD; Yuan-Teng Hsu[6*], PhD

[1]College of International Business, Zhejiang Yuexiu University, Zhejiang, China

[2]Shaoxing Key Laboratory of Intelligent Monitoring and Prevention of Smart City, Shaoxing, China

[3]Research Institute for Modern Economics and Management, Zhejiang Yuexiu University, Zhejiang, China

[4]College of International Finance and Trade, Zhejiang Yuexiu University, Zhejiang, China

[5]Center for Health Decision Science, Chongqing University of Posts and Telecommunications, Chongqing, China

[6]Research Center of Finance, Shanghai Business School, Shanghai, China

[*]all authors contributed equally

Corresponding Author:
Yuan-Teng Hsu, PhD
Research Center of Finance
Shanghai Business School
No 2271 West Zhong Shan Road
Shanghai, 200235
China
Phone: 86 21 64870020
Fax: 86 21 64288497
Email: yuanteng.hsu@gmail.com

Abstract

Background: Online health care services are a possible solution to alleviate the lack of medical resources in rural areas, and further understanding of the related medical service pricing system would contribute to improvement of the online health care community (OHC). Although many studies have investigated the OHC, the impact of physicians' reputations and wage levels on consulting prices in the OHC has rarely been discussed in the literature.

Objective: This study was designed to explore the determinants of consulting prices in the OHC. We addressed the following questions: (1) Are the prices of online health consultation services affected by wage levels at the doctor's location? (2) How does a physician's online and offline reputation affect their consulting prices?

Methods: Employing a large-scale sample of 16,008 doctors in China, we first used descriptive statistics to investigate the determinants of consulting prices in their entirety. Hierarchical linear modeling was then used to investigate the determinants of consulting prices in the OHC.

Results: The empirical results led to the conclusion that if doctors have more elevated clinic titles, work in higher-level hospitals, have better online reputations, and/or have made more past sales, their consulting prices will be higher. Additionally, the wage level in the city in which the doctor is working determines their opportunity cost and therefore also affects consulting prices.

Conclusions: The findings indicate that the characteristics of the doctor, the doctor's online reputation, and past sales affect the consulting price. In particular, the wage level in the city affects the price of the consultation. These findings highlight that the OHC is important because it can indeed break through geographical restrictions and give rural residents the opportunity to obtain medical service from doctors in big cities. However, doctors from cities often charge higher fees because of their higher opportunity cost. The results reveal that one of the most important functions of the OHC is to reduce the medical disparity between urban and rural areas; however, planners appear to ignore the possibility that rural residents with lower incomes may not be able to afford such high medical consultation costs. Therefore, the government should consider providing incentives to encourage urban doctors to provide discounts to rural residents or directly offer appropriate subsidies.

(J Med Internet Res 2021;23(7):e29170) doi:10.2196/29170

XSL•FO
RenderX

KEYWORDS

online health care industry; consulting pricing; reputation; wage level; hierarchical linear modeling; modeling; online consultation; pricing; linear regression; consultation; physician; eHealth

Introduction

Background

Health care is a fertile field for research in the service area, and is an expensive, complex, and universally used service that substantially affects everyone [1]. Currently, certain activities related to health care are becoming more convenient due to ubiquitous internet connectivity; these new services may be referred to as online health care or telemedicine. Online health care may be defined as the use of medical information remotely via electronic communication to deliver health care services or improve a patient's clinical condition. The online health care community (OHC) is growing quickly, which is changing the traditional channels and approaches to health care services [2]. The Telehealth Index: 2017 Consumer Survey indicated that 50 million Americans would prefer to switch their health care service provider to one that offered telemedicine, compared to 17 million in 2015. Additionally, approximately 60% of patients with chronic illnesses indicated that they would like to manage their conditions via remote diagnosis with their physicians regularly [3]. The popularity of the OHC in developed countries is unsurprising because it allows patients to save time and travel costs [4]. However, China's overall per capita medical resources are relatively inadequate compared with those of more developed countries. According to the 2018 Organisation for Economic Co-operation and Development health resource data, China has only 1.80 doctors per 1000 people, which is relatively few compared with the numbers in the United States (2.60), the United Kingdom (2.80), Germany (4.30), and Norway (4.80) [5]. Thus, the development of the OHC is crucial not only for the sake of convenience but it is also an important means to reduce pressure on the entire medical system. Specifically, the OHC can reduce the length of hospital waiting times as well as rural-urban health disparities [6,7]. Therefore, it is worthwhile to devote attention to topics related to the OHC.

The OHC has developed very rapidly in China [8,9]. Up to 194.8 million people used online health care services in December 2016, accounting for 26.6% of the total number of internet users [10]. For these users, the three types of services with the highest usage rates were medical information inquiry (10.8%), online registration (10.4%), and online consultation services (6.4%). The first two services are usually available for free or for a small fee; in China, where the family doctor system is not as widespread as it is in developed countries, this has made online consulting services more important. However, online consulting services require payment of fees based on the doctor's pricing. Additionally, online consulting prices are not regulated according to a certain pricing standard, and therefore differences in the prices charged by individual doctors can be very large. This raises an important question about whether the pricing is reasonable. Consumers want to be able to judge whether the consulting price is too high, while doctors want to set reasonable prices for consultation based on their personal conditions. In this study, we investigated the factors that might influence online health care consulting prices.

Two related points deserve further explanation. First, it is important to have a more complete understanding of the current status of China's OHC to better understand the characteristics and limitations of this study. Xie et al [8] reported that there are 43 internet hospitals in China, 41 of which have licensed medical qualifications and 34 of which provide online consulting services. In addition to traditional hospitals adding various new remote services for patients, some internet companies have also obtained the necessary legal qualifications to provide medical services online. These internet companies have the advantage of being free from geographical and institutional restrictions. Specifically, they can hire qualified doctors from a variety of traditional hospitals across the country to provide medical services during their free time. In this study, data were collected from a well-known internet company represented by the Haodf website (*hao* means "good" and *dai fu* means "doctor" in Chinese). After appropriate data filtering (see "Data Collection and Sample Characteristics" section in the Methods for a detailed description), the total number of doctors in our sample was 16,008, from 1628 traditional hospitals in 30 provinces. With these data, we carried out a broad investigation of the determinants of consulting prices in China's OHC.

Second, online health care consultation is substantially different from tangible products as it possesses certain characteristics of services, including intangibility, inseparability, variability, and perishability [11]. Although there has been substantial research on consulting fees [12-16], to the best of our knowledge, this study is the first to explore consulting fees in the OHC based on large-scale empirical data collection. Compared with traditional electronic market items, online health care consultation can be regarded as a typical credence good. When common products are purchased, customers can recognize their utility or quality after using them. In other words, people can distinguish whether the product is good or not. However, with credence goods, it is usually hard for customers to verify the quality of the product they have received [17-19]. For example, in the area of online health care consultation, when patients receive suggestions from the doctor, in general, it is hard for them to know for certain whether the doctor's advice is appropriate for their personal situation. Similarly, it is difficult for patients to judge whether the pricing of consulting services is reasonable. This may induce the problem of moral hazard [20,21], which implies that physicians do not bear the full responsibility for the costs or risks associated with their providing online health care consultation and so may make less of an effort to reduce these costs. Although this study cannot propose a complete solution to the problem of moral hazard, it can provide several key factors that could help to explain physicians' consulting prices. This would in turn help to reduce the occurrence of mispricing, either deliberately or through carelessness, which will ultimately contribute to further desirable developments in the OHC.

XSL•FO
RenderX

Research Problem

The purpose of this study was to identify the determinants of online health care consulting prices. Owing to restrictions of data accessibility, two categories of determinants were included in our analysis. The first, relevant to economic factors, is the provincial wage level. The other category, obtained from the publicly available contents on the Haodf website, is related to the reputations of online and offline doctors. Reputation is a key indicator linked to trust [22], and is recognized as one of the primary factors affecting consumer behavior and seller performance, especially in online markets [23,24]. In the context of the OHC, a higher degree of information asymmetry may exist because OHC products have the characteristics of credence goods and items in online markets. Therefore, the particular nature of health care makes it necessary to consider both offline and online reputations so as to reduce information asymmetry. Previous empirical studies have focused on the correlation between reputation and price [15,23]. However, these studies investigated the online and offline reputations separately, thus not associating the combined offline reputation and online reputation with price. Given the multiple levels of data structure, in this study, hierarchical linear modeling (HLM) [25,26] was used to verify the relationships between online health care consulting prices and these determinants.

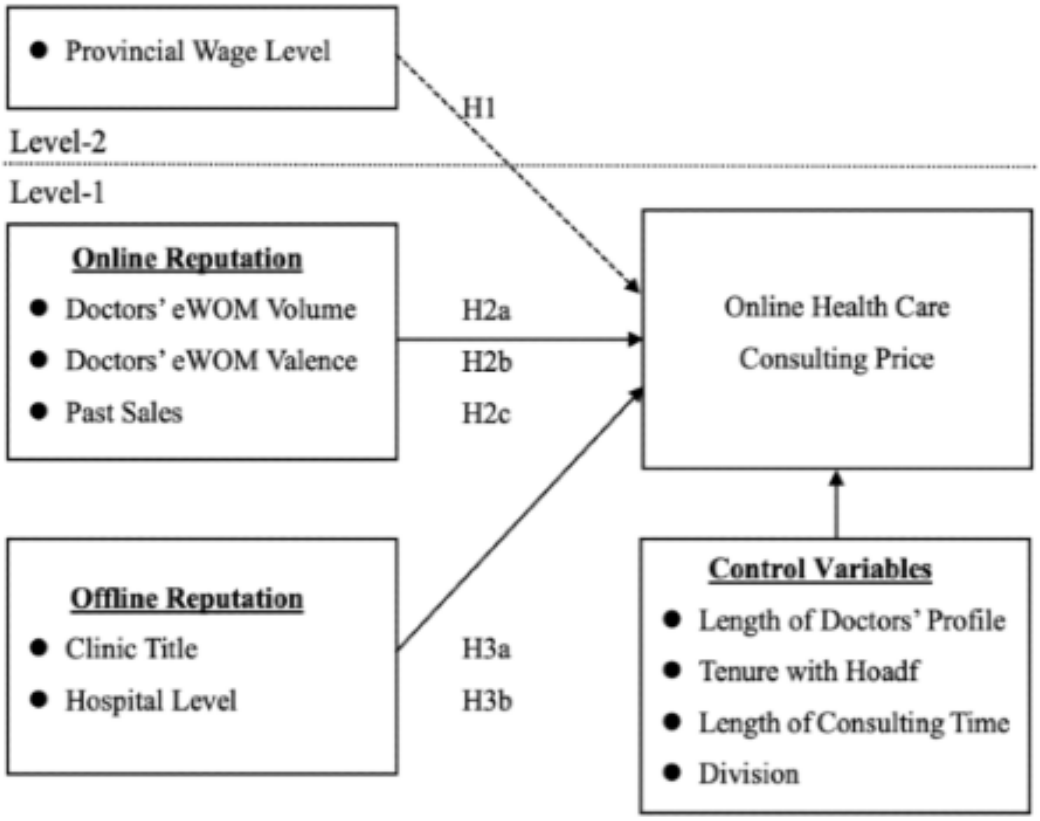
Methods

Research Model

Overview

We developed an explanatory research model for this study, as shown in Figure 1. Several hypotheses related to online health care consulting prices are proposed. An integrated multilevel model of online health care consulting prices, economic factors, and online and offline reputation is proposed. The dotted line separates the Level-2 variable and the Level-1 variables. The arrows crossing the dashed line represent cross-level relationships with the dependent variable. We explain how the data were analyzed in this multilevel model in further detail below. Additionally, four control variables were included in this research model.

Figure 1. Proposed research model. eWOM: electronic word of mouth.



Theoretical Background: Service Pricing Approaches

Among the many approaches to service pricing [13,27,28], cost- and competition-oriented pricing strategies have dominated the pricing of services [29]. The cost-oriented pricing approach arrives at a service price by considering all of the costs and adding a desirable profit margin [29,30]. In the competition-oriented pricing approach, the price is set to meet that of the competitors or to accord with the market situation [31,32]. Although the simplistic nature of both approaches brings advantages of usability and rapidity, to accurately reflect the practices of complex business organizations, many extended approaches also take the characteristics of the service into consideration. For professional service providers such as physicians, Hoffman and Arnold [33] proposed an extended cost-oriented pricing approach, which includes not only the traditional factors of fixed costs, variable costs, and the profit desired but also a service characteristic premium. In particular, they take the service characteristics premium to be based on essentiality, durability, and tangible intrinsic value-added, which are also widely regarded as service characteristics. Similarly,

XSL•FO
RenderX

Arnold et al [31] proposed a premium competition-oriented pricing approach in which the service price is equal to a differentiation premium plus the average competitor's price. They suggest that the differentiation premium could be a function of availability, reputation testability, commitment incentive, and price sensitivity premiums. However, in practice, the precise definition of the differentiation premium should be dependent on the specific issue under consideration. Recently, studies have explored the factors that influence consulting pricing [14-16]. McLachlin [14] suggested that consulting engagements are successful if the client is satisfied that the consultant meets client expectations, which may enhance the consultant's reputation and expectations of future revenue streams. These clients are more willing to pay a price premium for their services [34]. In addition, Momparler et al [15] showed that client satisfaction with the consulting team positively and strongly affects consulting fees. This satisfaction dimension may allow consultants to charge higher fees through the application of greater leverage while setting prices during the contracting process. Moreover, Lassala et al [16] performed a fuzzy-set qualitative comparative analysis to ascertain whether consulting-client satisfaction would explain differences in consulting fees and determine the conditions that lead to high consulting fees. Therefore, the aim of this study was to explore the critical determinants that affect online health care consulting prices, as this information can be used to improve the existing approaches to pricing. Specifically, these determinants can be regarded as components of the service characteristics premium or the differentiation premium in cost-oriented pricing or competition-oriented pricing, respectively.

We would like to further explain the importance of service characteristics for the analysis of online health care consulting prices by exploring differences between general services and those in the OHC. First, all service providers in the OHC are qualified doctors, and therefore this industry is professionally oriented; this suggests that service characteristics would be expected to play a critical role when providers price their services [33]. Second, almost all of the doctors in the OHC are online only part time, meaning that they provide online consultation during their nonworking hours. This service requires only a video or voice device and an internet connection; it is not necessary to rent an office or hire employees. Therefore, doctors have near-zero fixed costs, and their variable costs are closely related to their own opportunity costs. Third, all services are available online. Patients can find appropriate doctors according to information about specific doctors, without geographical restrictions. This implies that a doctor's individual characteristics (eg, clinic title or online rating score) become more important for patients wishing to make the best decision in choosing a doctor [23,35]. In the following sections, we introduce the three types of determinants used in this study: economic factors, online reputation, and offline reputation. We then present the complete research model.

Economic Factors

We first discuss the effects of an economic factor: the provincial wage level. This is the local salary level at the doctor's place of work. Intuitively, it seems obvious that a higher wage level usually means that doctors might experience higher opportunity costs when they provide online health care consulting services. As suggested by the cost-based approach [29,30], when doctors experience higher opportunity costs as a result of spending their time and effort offering a service, they will naturally request higher remuneration. Accordingly, we propose the following hypothesis:

> *Hypothesis 1: When the provincial wage level in the work area is higher, a doctor's online health care consulting price will be higher.*

Online Reputation Factors

Given that there are fewer repeated interactions between two transacting parties on the internet, Ba and Pavlou [23] pointed out that building appropriate feedback mechanisms can bring about a calculus-based credibility trust. There are three sources of trust in the business world: familiarity, calculativeness, and values [36]. In addition, two types of trust can be distinguished: benevolence and credibility [22,37]. In the context of an online transaction, Ba and Pavlou [23] suggested the term "calculus-based credibility trust" to express the relevant characteristics of this kind of market. Moreover, information asymmetry on the internet generally induces transaction-specific risks, but trust can mitigate these risks and thereby generate price premiums for reputable sellers. However, a major limitation in obtaining a sample from secondary data rather than using an experimental design is that it is not easy to measure trust perceptions with the former approach; therefore, some research such as the second study in Ba and Pavlou [23] has focused on evaluating the existence of a direct relationship between feedback mechanisms and price premiums. It is worthwhile to examine the meanings of the notions of feedback mechanisms and price premiums more carefully. First, in our context, to make it easier to mitigate the same problem of information asymmetry as can be found with other internet platforms, the Haodf website has provided a feedback mechanism that can produce reputational outcomes resulting from each agent's behavior. Second, following Ba and Pavlou [23], the price premium of one specific product is defined as the amount of money above the average price charged by different sellers. Because this study only considers one type of product, it is not necessary to subtract the average price to effectively compare the products at different price levels. Exploring the variation in consulting prices is essentially the same as examining the providers' price premiums. Therefore, higher consulting prices may be viewed as compensation to doctors for creating a good reputation, whereas lower consulting prices may be viewed as compensation to patients for bearing a greater transaction risk. As elaborated above, we believe that if the doctor has a better online reputation, their online consulting price will be higher. In the context of the OHC, the doctor's online reputation is based on patients' overall assessment of the doctor and indirectly reflects their level of trust in the doctor [23,38]. We will introduce three variables related to doctors' online reputations; the values of these variables can be determined based on information collected from the Haodf website.

The online reputation reflects a doctor's past actions, and the site allows patients to directly evaluate doctors and share their

XSL•FO
RenderX

assessments with others. In general, this type of online reputation can be measured through assessing a doctor's electronic word of mouth (eWOM) [39]. Two major attributes of eWOM, volume and valence [40,41], are considered in this study. A large volume of reviews makes a doctor stand out from the crowd, which might attract more attention from patients [40,41], and the valence (a positive or negative direction) of the patient reviews usually represents the quality of the doctor reviewed [42-44]. Moreover, an effective feedback mechanism, which can help buyers distinguish between sellers, can create price premiums [45]. In the buyer-and-seller relationship literature, Ba and Pavlou [23] found that there is a positive relationship between a seller's reputation and the price asked on eBay. This suggests that a better eWOM will come with a more favorable evaluation of a doctor and make patients more likely to pay higher prices. In addition, a doctor's past online sales partially reflect popularity, which is a signal of consulting quality [46-48]. Classical observational learning studies have found that customers tend to take cues from selections made by their peers because they infer the quality of products from such peer selection [49,50]. A higher past sales volume can also increase patients' trust and facilitate online transactions by reducing perceived risk. In this way, we expect that the past sales volume will positively affect the price of online health care consultations. Therefore, we propose the following hypotheses:

Hypothesis 2a: When a doctor has a larger eWOM volume, their online health care consulting price will be higher.

Hypothesis 2b: When a doctor has a more favorable eWOM valence, their online health care consulting price will be higher.

Hypothesis 2c: When the volume of a doctor's past online sales is larger, their online health care consulting price will be higher.

Offline Reputation Factors

Reputation is one of the important service characteristics [31]. A physician's offline reputation is a useful signal that can increase patient trust. For example, one patient's positive experience can increase trust in the doctor by other patients and minimize their perception of risk [51]. In the OHC context, a physician's offline reputation includes two aspects: (1) clinic title, which is the medical capability of the doctor assessed by the government agency based on the doctor's comprehensive ability; and (2) the level of the hospital with which the doctor is affiliated, which also reflects the doctor's ability. The clinic title is nationally standardized with four levels: resident physician, attending physician, associate physician, and chief physician (listed here from the most junior to the most senior level). Similarly, there are corresponding levels of the hospitals that doctors belong to: Level III, Level II, and Level I [52]. Level III is the highest level; hospitals at this level have more beds, better equipment, highly skilled doctors, and more comprehensive health care quality [53,54]. A doctor with a higher clinic title or coming from a high-level hospital usually translates to more experience, which leads to increased trust in the doctor, a situation that should reduce information asymmetry risks and result in a price premium for providers. Thus, we propose the following hypotheses:

Hypothesis 3a: When a doctor has a higher-level clinic title, their online health care consulting price will be higher.

Hypothesis 3b: When a doctor is from a higher-level hospital, their online health care consulting price will be higher.

Control Factors

Four control variables were included in this research model. First, we considered the length of the personal profile as benefits accrue to the patient if the information can be obtained without additional search costs. In general, if there are more words in the profile, then more details about the doctor, such as academic qualifications, clinical experience, area of expertise, or other capabilities, will be accessible. The added depth of information can increase a patient's confidence in making the decision to use the doctor's services [55]. Second, since the Haodf website has been established for more than 10 years, the length of time the doctor has had their profile activated on the website was also considered in our model. We refer to this variable as tenure with Haodf; tenure is generally used in studies related to the OHC [7,53,56]. Third, due to the fact that the length of time that is considered to constitute a consultation varies (eg, some doctors charge per 15 minutes and some charge per 10 minutes), we added the length of consulting time as a control variable to reduce its influence on our analysis. Finally, we also took the classification of the doctor's expertise into consideration and therefore added the doctor's division as another control variable. Given the complexity in assessing a doctor's expertise, we used the classification method provided by the Haodf website to divide the doctors into 10 divisions. Controlling this categorical variable was expected to reduce the effects of different types of expertise on online health care consulting prices.

Data Collection and Sample Characteristics

We collected a large amount of public data from the Haodf website [57] to explore our research questions. Founded in 2006, the Haodf website is the largest doctor-patient interaction platform in China and is considered to be the most professional and trustworthy OHC. More than 490,000 physicians from 7500 hospitals nationwide are included on this website, with 145,000 physicians having completed real-name registration on the website as of 2017. In addition, we chose this OHC for two reasons. First, online doctor consultation services on Haodf have involved charges since the middle of 2016. Through the Haodf website, patients can consult with their doctor via text, voice, or video as needed regardless of time and geographical restrictions. These services are offered to patients for different fees. Since telephone consultation services were the most widely used mode during the data collection period, this study mainly selected the telephone consulting price for discussion. We can also observe the telephone consulting price per doctor directly on the website. Second, since Haodf is one of the most popular OHCs in China, many recent studies have used data sources from this website to explore various research questions [7,52,53,56,58,59], which means that the data collected from

XSL•FO
RenderX

this site are reliable and representative. Our data collection procedure is described in further detail below.

We adopted web crawler technology for the purpose of exploring more than 423,000 doctors' profiles on the Haodf website [57]. Within the sample, not every doctor provided online consulting services. Specifically, only 44,780 of the 423,000 doctors offered this kind of service and announced the related prices on their personal pages. The website is designed such that when a patient pays for a service, one review record will be added to the doctor's page, even if the patient does not comment. Thus, we can observe the number of past sales through the number of reviews on each doctor's page. To ensure that real tradable prices were used in the analysis, the final sample considered only 16,008 doctors who had actually provided at least one online consulting service to a patient. Data from the Haodf website have been used in many recent studies [53,56]; therefore, readers interested in obtaining more information on the Haodf website can refer to this literature. Additionally, for convenience, we also provide screenshots relevant to this kind of service on the Haodf website in Figure 2 to illustrate the form in which the doctor's basic information (eg, name, position, and division), details about the service process, the consulting price, and patients' reviews can be found.

Figure 2. Haodf website screenshot. WOM: word of mouth; WVOL: volume of electronic word of mouth.

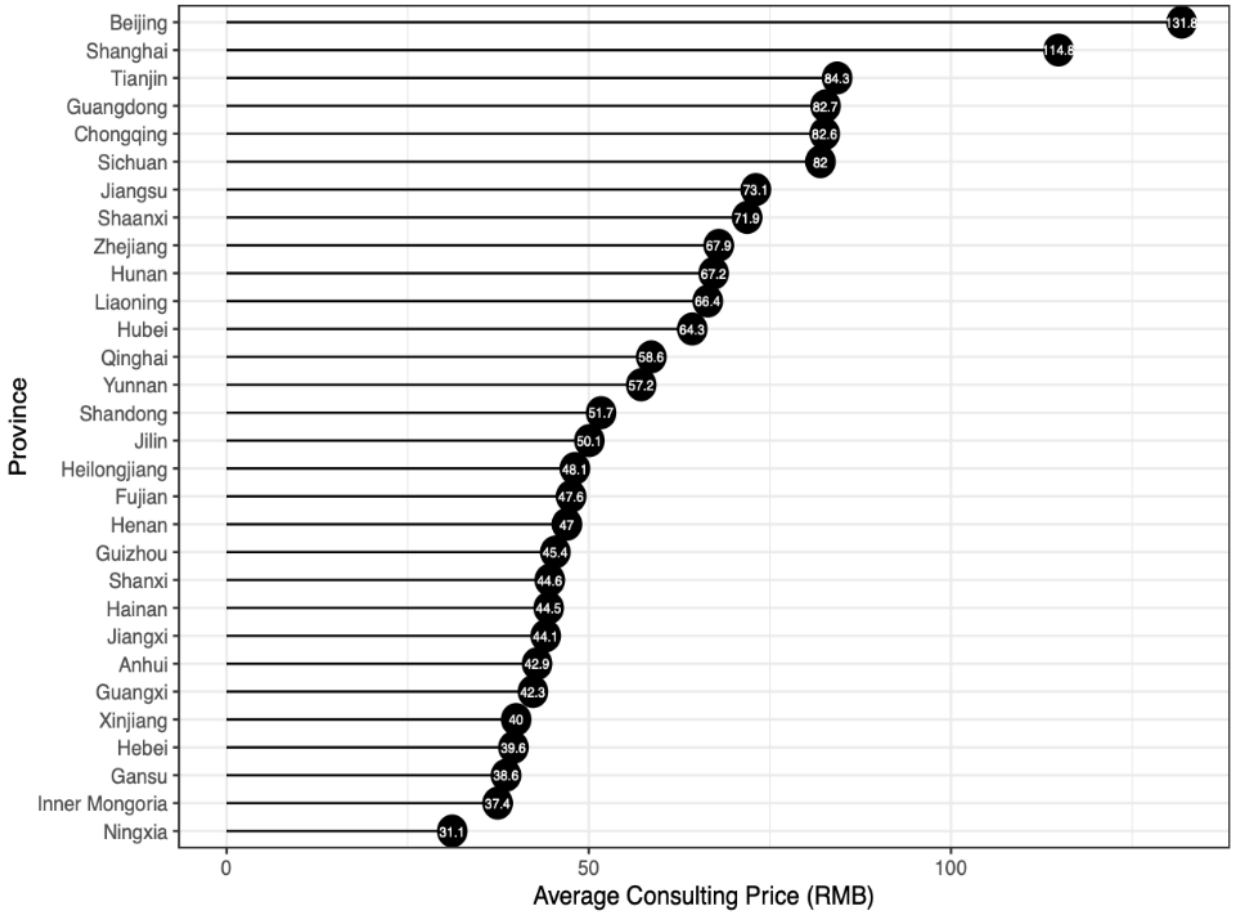


Certain characteristics of the sample deserve further mention. First, the total past sales, or the number of patients who had paid for a consulting service on the Haodf website, was 320,265. In other words, many people had used this type of service, which also reflects the degree to which the OHC has been accepted in China. Second, these 16,008 doctors came from 1628 different hospitals, representing 10 different kinds of professional divisions, and were from 30 different provinces in China. This shows that our sample was not confined to a particular organization, medical specialty, or geographical area. Third,

XSL•FO
RenderX

this online health care consulting service charged on average 84.7 yuan (about US $13; 1 yuan=~US $0.15) per consultation, or 7.59 yuan per minute. Notably, the consulting price was clearly affected by the province in which the doctor's hospital was located, as shown in Figure 3. For example, the average consulting price in Beijing was nearly double that in more than half of the other provinces. To take this difference into account, the provincial wage level was included in our model to make the differences in online consulting prices in the various geographical areas more apparent.

Figure 3. Online health care consulting prices in different provinces.



Note: 1 RMB is being exchanged for about 0.1548 USD on June 19, 2021.

Measures

Dependent Variable: Consulting Price

The dependent variable in this study was the telephone consulting price, which is the price plan set by doctors who provide online health care telephone consulting services. For example, a doctor might charge 200 yuan per 10 minutes or 500 yuan per 15 minutes. The patient can then make an appointment with the doctor directly and pay the relevant fees online. It is important to note that the patient always pays for a block of time instead of paying for the precise amount of time spent in consultation. In other words, when the patient chooses a price plan such as 500 yuan per 10 minutes, they will be charged 500 yuan even if the actual consulting time is less than 10 minutes. To guarantee the fairness of the transaction, the Haodf website acts as a third-party impartial institution. At the appointed time, the patient makes a call to the Haodf website and an employee assists them in contacting the doctor for the telephone consultation. After completing the telephone consultation, the patient writes a comment about the service that has been provided. If the patient does not write a review, the system generates a record with an empty comment. In light of these facts, we defined the telephone consulting price as the listed price (in yuan) per instance of the online health care consulting service provided by a doctor, and the listed length of consulting time was added to the model as a control variable. In these terms, among the 16,008 doctors, approximately 70% charged a price less than 100 yuan, but 116 doctors charged an amount more than or equal to 500 yuan. In addition, the mean price was 84.7 yuan, but the median was only 57.5 yuan. These two results indicate that the distribution of online health care consulting prices has a significantly positive skewness; therefore, we used the online health care telephone consulting price in the natural logarithmic form in our analysis model.

XSL•FO
RenderX

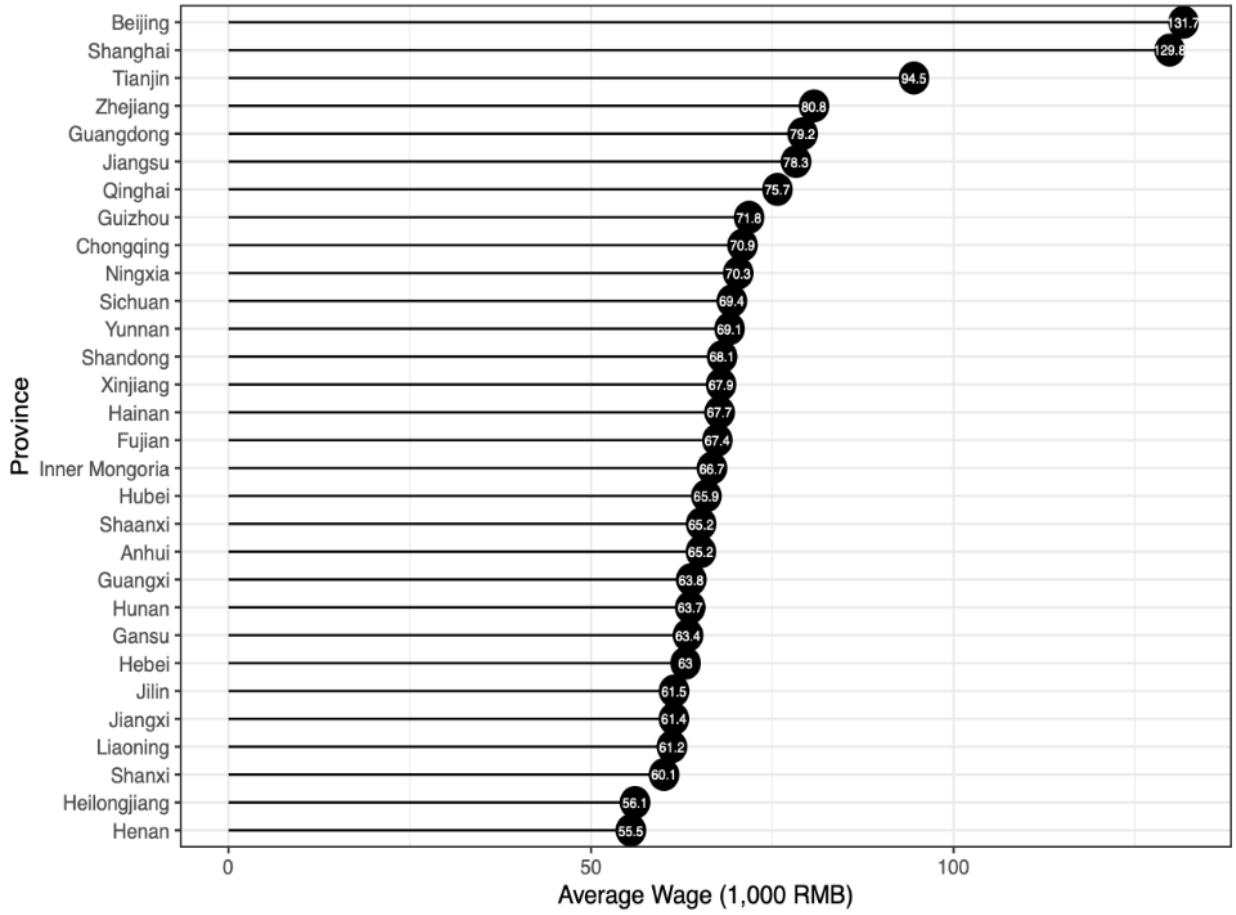
Independent Variables

Wage Level

The doctor's wage level was measured using the provincial average currency wage for employees in 2017. All doctors from the same province in the sample were related to the same provincial wage level. Therefore, the provincial wage level was regarded as a level-2 or group-level variable. We obtained the provincial wage level from the Wind Economic Database [60] and formally used the variable WAGE for representation. Figure 4 shows the wage levels in 30 provinces; these figures also indicate the significant relationship between the telephone consulting prices and the provincial wage level. For example, Beijing, Shanghai, and Tianjin are not only the top three cities in terms of average telephone consulting prices but are also the top three in terms of the provincial wage level.

Besides the level-2 variable, our model further included five level-1 or individual-level variables: eWOM volume, eWOM valence, past sales, clinic title, and academic rank.

Figure 4. Wage levels in different provinces.



Note: 1 RMB is being exchanged for about 0.1548 USD on June 19, 2021

eWOM Volume

The doctor's online reputation was measured using the volume of eWOM (WVOL) calculated by the number of reviews of the doctor. Because the distribution of the variable WVOL had significantly positive skewness, we used the natural logarithmic in our analysis model.

eWOM Valence

The doctor's online reputation was used to measure their overall score on the website. Specifically, after a patient receives a doctor's service, they can indicate that the doctor has a certain professional ability to evaluate the doctor, giving the doctor a score of 1 to 5 (5 being the best). The eWOM valence (WVAL) was then measured as the mean of the overall ratings of doctors.

Past Sales

The doctor's online reputation was also used to measure the past sales of their telephone consultation services. Even if the review content is empty, the system keeps all records of online health care consulting reviews created when users have finished receiving a service. Hence, based on the number of these reviews, we defined the variable PSALE as the number of patients who paid for the online health care services provided by doctors.

Clinic Title

The doctor's offline reputation was measured using the doctor's clinic title, which is the medical ranking of the doctor as evaluated by the government according to the doctor's comprehensive abilities. In China, the doctor's clinic title can

XSL•FO
RenderX

be specified as chief physician, associate chief physician, attending physician, or resident physician. We defined the clinic title as a dummy variable (CL), in which the positions of chief physician or associate chief physician are coded 1 and others are coded zero [7].

Hospital Level

The doctor's offline reputation was measured by the level of hospital to which they belonged. In this study, we defined hospital level as a dummy variable (HL), with 1 assigned to a doctor from a tertiary hospital and 0 otherwise [7].

Control Variables

We employed four control variables in this study. First, doctors can provide their personal profile information such as education, experience, and expertise for public reference on the Haodf website. We calculated the number of words in each profile to represent the level of individual information disclosure, which represented the value for the variable WORD. Second, because the Haodf website had been established for more than 10 years, how long the doctor had been a member might have affected their behavior on the website. To control for time effects, the TENURE variable was defined as the tenure with the Haodf website [56], which was measured as the number of days the doctor had their own profile activated on the website. Third, because the amount of time considered to constitute a single consultation was different for different doctors, it was necessary to take this difference into consideration. Therefore, LCT was defined as the length of a single consulting time. Finally, doctors have different major specialty areas such as surgery, internal medicine, pediatrics, and Chinese medicine. To reduce the effect of the characteristics of the doctor's specialty, we further added another control variable, DIV, which is a categorical variable with 10 major divisions, based on information reported on the website. The detailed variable definitions and measurements are shown in Table 1. It is important to note that the distributions of several variables were positively skewed; for example, approximately 80% of doctors had fewer than 500 words in their profile, but nearly 5% of doctors had more than 1000 words. To alleviate the problem of data being positively skewed in our analysis, the variable was transformed to the natural logarithm form and the letter "L" was added at the beginning of its code (eg, LWORD is the variable WORD in the natural logarithmic form).

Table 1. Variable measurements and sources.

Code	Variable	Measurements	Source
PRICE	Consulting price	The listed price of the online health care consulting service provided by doctors	Haodf
WAGE	Provincial wage level	The provincial average currency wage of employees in 2017	Wind Economic Database [60]
WVOL	Doctor's eWOM[a] volume	The number of reviews provided by patients indicating the doctor's professional ability on the website	Haodf
WVAL	Doctor's eWOM valence	Mean of the overall ratings in patients' reviews of the doctor	Haodf
PSALE	Past online sales	Number of patients who had paid for the doctor's online health care service	Haodf
CL	Clinic title	Clinic title of the doctor, dummy variable: CL=1 if the doctor's position was chief physician or associate chief physician; 0 otherwise	Haodf
HL	Hospital level	The level of hospital to which the doctor was affiliated, dummy variable: HL=1 if the doctor was from a tertiary hospital; 0 otherwise	Haodf
WORD	Length of doctor's profile	Number of words in the doctor's personal profile	Haodf
TENURE	Tenure with Haodf	The doctor's tenure with the Haodf website (in days), calculated based on the data download date minus the doctor's registration date on the website	Haodf
LCT	Length of consulting time	The listed length of one occasion of the consulting service (in minutes) provided by doctors	Haodf
DIV	Division	The doctor's division as categorized by the website, including internal medicine, surgery, gynecology/obstetrics, pediatrics, orthopedics, ophthalmology, oral health, cancer, Chinese medicine, and others	Haodf

[a]eWOM: electronic word of mouth.

Statistical Analysis: HLM

The hierarchical level of grouped data is a common consideration in many research contexts [61]. For example, to explore how job satisfaction influences work performance, different employees' individual characteristics are usually taken into consideration. In addition, since workers are part of different teams, the cohesiveness of each team might also be an important factor in the model. However, using simple linear regression techniques to perform such analyses may be insufficient given the lack of consideration to shared variance. To overcome this problem, HLM is one of the most popular methods that can simultaneously handle the relationships among variables both at the within-group and between-group levels [25,26]. In this study, the sample involved a hierarchy with two levels. Specifically, the higher level of the hierarchy (level 2) includes the provincial wage, and a lower level of the hierarchy (level 1) includes other variables (eg, consulting price, clinic title, or academic rank). All level-1 variables are nested within level-2 groups and have a share in the common impact of the level-2

XSL•FO
RenderX

variable. We therefore performed HLM as our statistical analysis, which explicitly takes into account this cross-level data structure. We implemented the related HLM analyses through the multilevel and nlme packages in R.

To examine whether our data met the prerequisites for HLM analysis, we adapted a one-way analysis to identify the within-group and between-group variance in the dependent variable. Formally, the relevant null model is as follows:

Level 1: $LPRICE_{ij}=\beta_{0j}+r_{ij}$ **(1)**

Level 2: $\beta_{0j}=\gamma_{00}+u_{0j}$

where $LPRICE_{ij}$ is the natural logarithm form of the consulting price measured for doctor *i* in province *j*, and β_{0j} is the mean of *LPRICE* for province *j*. In the combined form, the model is $LPRICE_{ij}=\gamma_{00}+u_{0j}+r_{ij.}$ In this way, the null model essentially means that the dependent variable is a function of a common intercept γ_{00} and there are two error terms: the between-groups error term u_{0j} and the within-group error term r_{ij}. In addition, we denote the variances r_{ij} and u_{0j} as τ_{00} and σ^2, respectively. Based on the general assumption of HLM, $cov(r_{ij}, u_{0j})=0$, the variance of LPRICE can be partitioned into between-group variance (τ_{00}) and within-group variance (σ^2). We can then calculate the intraclass correlation coefficient (ICC), which is the ratio of between-group variance to total variance, as follows:

$ICC=\tau_{00}/\tau_{00}+\sigma^2$ **(2)**

Once significant between-group variance in the dependent variable was determined, two types of HLMs were used to test the hypotheses of this study. First, we considered the random intercept model as follows:



Level 2: $\beta_{0j}=\gamma_{00}+\gamma_{01}LWAGE_j+u_{0j}$

$\beta_{kj}=\gamma_{k0}$, k=1, 2,…,5 **(3)**

where $LPRICE_{ij}$ is the natural logarithm form of the consulting price measured for doctor *i* in province *j*, and β_{0j} is the mean of LPRICE for province *j*. In addition, LWVOL, WVAL, LPSALE, CL, and HL are independent variables, and $X_{ij}^{(m)}$ represents the control variables, including LWORD, LTENURE, LCT, and DIV. The relevant definitions of the variables are shown in Table 1. We added the letter "L" at the beginning of the variable code to identify a variable in natural logarithm form. To further clarify the meaning of the random intercept model, we also rewrote it as the combined form:



where γ_{00}, u_{oj}, and r_{ij} *represent* a common intercept, the between-groups error term, and the within-group error term, respectively. Compared to the traditional ordinary least-squares regression model, the random intercept model has additional between-groups error terms, which allows the intercepts to be different among groups. However, the impact of independent variables on the dependent variable is fixed. In other words, for one specific independent variable, the coefficients or slopes (β_{kj}) in all groups (provinces) are the same. The second type of HLM, described below, will relax the fixed effect of β_{kj}.

The second HLM used in our analysis was the random coefficient model. The setup of level 1 in the random coefficient model is the same as that in the random intercept model, but at level 2 it involves an extra error term in β_{kj}. This model can be written as follows:



Level 2: $\beta_{0j}=\gamma_{00}+\gamma_{01}LWAGE_j+u_{0j}$

$\beta_{kj}=\gamma_{k0}+u_{kj}$, k=1, 2,…,5 **(4)**

where u_{kj} is the error term and other symbols have the same definitions as in the random intercept model. The combined form of the random coefficient model is:



Therefore, the random coefficient model has a more complex structure of error terms compared to the random intercept model, which allows not only the intercepts but also the coefficients of each independent variable to be different among various groups. Both the random intercept and random coefficient models were used in our empirical study.

Results

Descriptive Statistics

The total number of samples used in this study was 16,008, and the relevant descriptive statistics (means, SDs, minima, maxima) for the sample are shown in Table 2. Note that natural logarithm transformation was used to mitigate the issue of skewed distribution. In the following discussions of the HLM, several variables, including PRICE, WAGE, WVOL, PSALE, WORD, and TENURE, will be presented on the transformed scale, and others will be presented on the original scale. In addition, since CL and HL are dummy variables, their means represent the proportion of variable values equal to 1 for the entire sample. Overall, 72.3% of the doctors in our sample were chief physicians or associate chief physicians, and 90.2% of the doctors were from tertiary hospitals.

XSL•FO
RenderX

Table 2. Descriptive statistics of variables (N=16,008).

Variable	Mean (SD)	Minimum	Maximum
Consulting price (PRICE)	84.699 (82.257)	1	1500
Provincial wage level (WAGE)	90,961.600 (29,497.110)	55,495	131,700
Doctor's eWOM[a] volume (WVOL)	62.884 (104.995)	0	1,746
Doctor's eWOM valence (WVAL)	3.997 (0.349)	2.8	5
Past online sales (PSALE)	20.007 (98.232)	1	4062
Clinic title (CL)	0.723 (0.448)	0	1
Hospital level (HL)	0.902 (0.298)	0	1
Length of doctor's profile (WORD)	381.407 (696.944)	41	27,301
Tenure with Haodf (TENURE)	1780.000 (988.839)	12	3402
Length of consulting time (LCT)	11.121 (2.763)	5	30

[a]eWOM: electronic word of mouth.

The correlations between major variables are listed in Table 3. All correlation coefficients were significantly positive at the 0.1% level, except for the correlation coefficient between the doctor's hospital level and the length of the doctor's profile, which was not significant.

We first verified whether the hypothesized determinants of the online health care consulting price are truly influential. The doctors were divided into two groups based on the median for the four continuous variables or the category for the two dummy variables. We then determined whether the average price of the two groups was the same. For example, since the median of WVOL was 27, the full sample was divided into two groups: those with higher WVOL and those with lower WVOL. The average PRICE for the higher group was 109.9, whereas that for the lower group was only 59.6. Table 4 shows the relevant results for all determinants. Moreover, we also adopted two-sample *t* tests to verify the significance of the difference of means. In each case, the means of the two groups were not equal at the 0.1% significance level. Therefore, these preliminary findings revealed that a doctor with a higher offline reputation (whether in terms of clinic title or hospital level), receiving a greater eWOM volume, a higher eWOM rating, having made more sales in the past, or coming from a higher-wage province generally leads to their claiming a higher online health care consulting price. In the following sections, we provide more empirical evidence to support the argument that these determinants significantly affect the online health care consulting price, and how they affect the price.

XSL•FO
RenderX

Table 3. Correlation analysis

Variable	LPRICE[a]	LWAGE[b]	LWVOL[c]	WVAL[d]	LPSALE[e]	CL[f]	HL[g]	LWORD[h]	LTENURE[i]	LCT[j]
LPRICE										
r	1.000	0.388	0.461	0.429	0.489	0.359	0.202	0.353	0.286	0.390
P value	—[k]	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
LWAGE										
r	0.388	1.000	0.163	0.162	0.290	0.069	0.038	0.147	0.166	0.151
P value	<.001	—	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
LWVOL										
r	0.461	0.163	1.000	0.722	0.588	0.252	0.123	0.330	0.320	0.138
P value	<.001	<.001	—	<.001	<.001	<.001	<.001	<.001	<.001	<.001
WVAL										
r	0.429	0.162	0.722	1.000	0.419	0.180	0.230	0.258	0.134	0.100
P value	<.001	<.001	<.001	—	<.001	<.001	<.001	<.001	<.001	<.001
LPSALE										
r	0.489	0.290	0.558	0.419	1.000	0.195	0.092	0.254	0.350	0.252
P value	<.001	<.001	<.001	<.001	—	<.001	<.001	<.001	<.001	<.001
CL										
r	0.359	0.069	0.252	0.180	0.195	1.000	0.076	0.480	0.382	0.083
P value	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001	<.001	<.001
HL										
r	0.202	0.038	0.123	0.230	0.092	0.076	1.000	0.096	0.093	0.009
P value	<.001	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001	.24
LWORD										
r	0.353	0.147	0.330	0.258	0.254	0.480	0.096	1.000	0.398	0.107
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	—	<.001	<.001
LTENURE										
r	0.286	0.166	0.320	0.134	0.350	0.382	0.093	0.398	1.000	0.109
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	—	<.001
LCT										
r	0.390	0.151	0.138	0.100	0.252	0.083	0.009	0.107	0.109	1.000
P value	<.001	<.001	<.001	<.001	<.001	<.001	.24	<.001	<.001	—

[a]LPRICE: natural logarithm of consulting price.

[b]LWAGE: natural logarithm of provincial wage level.

[c]LWVOL: natural logarithm of the doctor's electronic word of mouth volume.

[d]WVAL: doctor's electronic word of mouth valence.

[e]LPSALE: natural logarithm of past online sales.

[f]CL: clinic title.

[g]HL: hospital level.

[h]LWORD: natural logarithm of the length of the doctor's profile.

[i]LTENURE: natural logarithm of the tenure with Haodf.

[j]LCT: length of consulting time.

[k]Not applicable.

XSL•FO
RenderX

Table 4. Results of two-sample t tests.

Variable	Mean (SD)	*t* value	*P* value
Clinic title (CL)		43.395	<.001
Low	49.324 (52.038)		
High	98.269 (87.516)		
Hospital level (HL)		17.618	<.001
Low	53.666 (72.607)		
High	88.088 (82.542)		
Doctor's eWOM[a] volume (WVOL)		40.670	<.001
Low	59.552 (63.915)		
High	109.922 (90.475)		
Doctor's eWOM valence (WVAL)		34.004	<.001
Low	66.630 (69.581)		
High	112.881 (92.010)		
Past online sales (PSALE)		46.096	<.001
Low	57.502 (57.557)		
High	115.244 (94.243)		
Provincial wage level (WAGE)		38.687	<.001
Low	62.776 (60.232)		
High	114.622 (97.464)		

[a]eWOM: electronic word of mouth.

HLM Analysis

Null Model

We hypothesized that both individual- and province-level variables would be significantly correlated with consulting fees. For these hypotheses to be supported, there had to be significant between-province variance in consulting fees. Therefore, we used HLM to estimate a null model as in Eq (1), in which no independent variables are specified for either the level-1 or level-2 function, to test the significance level of the between-groups variance in the dependent variable by examining the significance level of and the ICC. In our case, =0.108 (P<.001); thus, we rejected the null hypothesis (τ_{00}=0), which implies a significant difference of LPRICE between different provinces. In addition, the ICC was calculated to be 0.142 according to Eq (2), which means that there is a 14.2% difference in the consulting fee between provinces, whereas there is an 85.8% difference within provinces. According to the criteria recommended by James [62] and Cohen [63], the impact of between-groups variance should not be ignored in the context of our study. In other words, our data met the prerequisites for performing HLM analysis rather than the traditional regression models. Therefore, we proceeded to test our hypotheses using HLM.

As suggested by Hofmann and Gavin [64], for each estimated HLM model (described below), all independent variables at level 1 were grand mean–centered. In addition, we defined $R^2_{within\text{-}group}$ as the proportion of within-group variance explained by the specific model compared with the null model and $R^2_{between\text{-}groups}$ as the proportion of between-group variance explained by the specific model compared with the null model. Moreover, following the formula reported in Raudenbush and Bryk [26], we further calculated the total variance explained by the dependent variable as follows:

$$R^2_{total}=R^2_{within\text{-}group}\times(1-ICC)+R^2_{between\text{-}groups}\times ICC \quad \textbf{(5)}$$

where ICC is defined as in Eq (2).

Random Intercept Model

Table 5 presents the HLM analysis results, including the coefficients, standard errors, *t* values, and *P* values, for the two models with the sample of 16,008 physicians. In the random intercept model as in Eq (3), the results showed a positive relationship between doctors' wage levels and online health care consulting price, thereby supporting H1 ($R^2_{within\text{-}group}$=0.431, $R^2_{between\text{-}groups}$=0.769, R^2_{total}=0.479). Thus, when a doctor comes from a province with a higher wage, they generally charge a higher price for online health care consulting. We also found that the online reputation of doctors indicated by variables such as eWOM volume, eWOM valence, and past sales had positive associations with the online health care consulting price, as represented by the corresponding coefficients, which in turn supports H2a, H2b, and H2c. In other words, when a doctor receives more WOM volume, has a higher eWOM rating, and has had more sales in the past, they charge a higher consulting fee. Finally, a significant positive relationship between offline

XSL•FO
RenderX

reputation (such as clinic title and hospital level) and consulting price was found, which supports H3a and H3b, suggesting that consulting price results from higher clinic titles or a higher-level hospital. All the estimates were statistically significant (P<.001).

Table 5. Hierarchical linear model results for the effects of determinants on online health care consulting prices.

Variable	Random intercept model			Random coefficient model		
	Coefficient[a] (SE)	*t* value[b]	*P* value	Coefficient[a]	*t* value	*P* value
Level-2 variable: LWAGE[c]	0.751 (0.149)	5.050	<.001	0.730 (0.067)	10.893	<.001
Level-1 variables						
Intercept	−0.537 (0.048)	–11.203	<.001	−0.619 (0.036)	–17.230	<.001
LWVOL[d]	0.060 (0.006)	9.876	<.001	0.069 (0.012)	5.598	<.001
WVAL[e]	0.412 (0.022)	18.551	<.001	0.371 (0.037)	10.027	<.001
LPSALE[f]	0.091 (0.004)	20.566	<.001	0.073 (0.008)	8.805	<.001
CL[g]	0.378 (0.013)	29.304	<.001	0.411 (0.017)	24.345	<.001
HL[h]	0.325 (0.017)	19.207	<.001	0.320 (0.017)	18.996	<.001
Control Variables						
LWORD[i]	0.079 (0.006)	12.523	<.001	0.071 (0.006)	11.056	<.001
LTENURE[j]	0.017 (0.008)	2.275	.02	0.013 (0.008)	1.766	.08
LCT[k]	0.082 (0.002)	45.378	<.001	0.082 (0.002)	45.624	<.001
DIV[l]: Surgery	−0.128 (0.017)	–7.715	<.001	−0.131 (0.017)	–7.921	<.001
DIV: Gynecology-obstetrics	0.059 (0.021)	2.745	.006	0.062 (0.021)	2.881	.004
DIV: Pediatrics	0.049 (0.020)	2.490	.01	0.055 (0.020)	2.778	.06
DIV: Orthopedics	−0.211 (0.023)	–9.121	<.001	−0.212 (0.023)	–9.187	<.001
DIV: Ophthalmology	−0.018 (0.026)	–0.694	.49	−0.017 (0.026)	–0.651	.52
DIV: Oral health	−0.189 (0.031)	–6.021	<.001	−0.191 (0.031)	–6.118	<.001
DIV: Cancer	−0.047 (0.030)	–1.585	.11	−0.042 (0.029)	–1.424	.15
DIV: Chinese medicine	−0.345 (0.030)	–16.375	<.001	−0.337 (0.021)	–16.023	<.001
DIV: Others	0.024 (0.017)	1.394	.16	0.026 (0.017)	1.510	.13

[a]Standardized regression coefficient.

[b]Degrees of freedom=28 for provincial wage level and 15,961 for the other variables.

[c]LWAGE: natural logarithm of provincial wage level.

[d]LWVOL: natural logarithm of doctor's electronic word of mouth volume.

[e]WVAL: doctor's electronic word of mouth valence.

[f]LPSALE: natural logarithm of past online sales.

[g]CL: clinic title.

[h]HL: hospital level.

[i]LWORD: natural logarithm of the length of the doctor's profile.

[j]LTENURE: natural logarithm of the doctor's tenure on Haodf.

[k]LCT: length of consulting time.

[l]DIV: division.

Random Coefficient Model

Even when adopting the random coefficient model, which allows the coefficients in all groups to be different, the main characteristics of these results also showed positive associations with the online health care consulting price (Table 5; $R^2_{within\text{-}groups}$=0.437, $R^2_{between\text{-}groups}$=0.730). In comparison with the null model, the R^2_{total} of the random intercept and random coefficient models were both exactly 0.479, implying that both models represent an obvious improvement over the null model and better explain the online health care consulting price. Additionally, the effects of control variables, LWORD and LCT, were significantly positive and were all consistent in both models. However, the effect of LTENURE was significantly

XSL•FO
RenderX

positive only in the random intercept model. In summary, based on the above outcomes, all of the proposed hypotheses were clearly supported. We will provide further evidence supporting the robustness of our results in the next subsection.

Robustness Tests for the Subsample

To test for robustness, a subsample was selected based on two different criteria. First, it is necessary to certify the authenticity of the observed online health care consulting prices. As mentioned above, the sample included only doctors who had completed at least one consultation at their own requested price. Nevertheless, it might be argued that the number of transactions is too small, which would make the online health care consulting price not suitable as a reference point. Thus, we verified whether the HLM results were robust for doctors with at least 10 past sales. Specifically, we ignored relatively inactive physicians, resulting in a reduction of the sample size to 4055. Regarding this subsample, since the random intercept and random coefficient models produced similar results in all robustness tests, we show only the results of the random coefficient model in Multimedia Appendix 1. The coefficients were substantially similar to those presented in Table 5. Second, we further examined the robustness of different subperiods. Specifically, we divided all of the doctors into two subsamples based on whether they had joined the Haodf website before or after January 1, 2014. We chose this date as the department sample for two main reasons. First, this point was chosen to better understand whether the length of time as a member on the platform would produce consistent results for our proposed model. Second, since the Haodf platform provided more complete mobile apps after 2014, we believe that doctors who joined the platform after this date may be more inclined to use a phone to communicate with their patients. Two subsamples were used: Subsample 2 and Subsample 3. Multimedia Appendix 1 shows that all independent variables still had significantly positive values. In summary, the additional empirical evidence provided here further confirmed the robustness of the relationship between the level-2 and level-1 variables and online health care consulting prices.

Discussion

Principal Results

The development of the OHC has become increasingly important, especially in countries where there are relatively few medical resources such as China [7,8]. A common type of service in the OHC is online health care consultation, which allows patients and doctors to make contact and exchange medical information remotely [59,65]. Since reasonable and stable prices are an important factor in ensuring the sustainable development of any industry, we have focused on investigating the determinants of online health care consulting prices. To the best of our knowledge, this study is the first to target this topic through analysis of large-scale empirical data. The main theoretical contribution of this study is that of enhancing and complementing the past service pricing literature to make it more applicable to the OHC. Specifically, the consideration of both cost-based and competition-based pricing approaches involves service characteristics [31,33]; however, there are many differences between the features of online health care consultation and general services, such as more specialized service providers, zero fixed costs, and online service in online health care consultation. Thus, we propose six determinants of online health care consulting prices to bridge this gap in knowledge of this field. Our empirical results support the conclusion that working area wage levels, eWOM volume, eWOM valence, past sales, the doctor's clinic title, and the level of the doctor's affiliated hospital have significantly positive effects on the doctor's online health care consulting price. Moreover, we suggest some corresponding management implications.

From the patients' perspective, since doctors might exploit informational asymmetry to defraud them [17], patients hope to avoid being subjected to an overcharge or low-quality service. We suggest that in practice, patients can first determine a basic price that is calculated based on the average price of other doctors with a similar specialty and then further adjust their expected price through our proposed determinants. For instance, when the price requested by the doctor is higher than the basic price, patients can check if this doctor is a chief physician, the eWOM is better, the number of past sales is larger, or the doctor is from a big city. From the opposite perspective, doctors can set reasonable or acceptable prices after reviewing their personal conditions based on certain proposed determinants. In particular, similar to the results in previous studies, our results also verified that eWOM valence and volume have positive effects on prices [43,44]. Therefore, doctors who want to offer online health care consulting services should manage their online reputation well. Companies providing online health care services, such as the Haodf website, should construct an effective feedback machine that can increase the patient's trust or reduce the uncertainty of online transactions [23]. Eliminating information asymmetry is definitely one of the pressing tasks in developing the OHC industry. Finally, our findings may be of interest to government authorities who wish to develop the OHC. There is no doubt that online health care services can indeed break through geographical limits and give rural residents the opportunity to obtain medical service from big cities. However, doctors from cities often charge higher fees because of the higher opportunity cost. One of the most important functions of the OHC is to reduce the medical disparity between urban and rural areas; however, planners seem to ignore the fact that rural residents with lower incomes may not be able to afford such high medical consultation costs. Therefore, the government can provide incentives to encourage urban doctors to give some discounts to rural residents or can directly provide appropriate subsidies. We believe that specific development of the OHC should not be undertaken simply for the sake of convenience. More importantly, it should be the case that people who truly need medical resources can receive effective medical care immediately.

Limitations and Future Work

There are some limitations to this study as well as indications of possible directions for future work. First, a cross-sectional study design was used for data collection in this study. The sample might have led to results that are time-sensitive. For example, diseases prevalent in different seasons can affect the

XSL•FO
RenderX

behavior of physicians. In addition, the Haodf platform regularly provides discounted prices to a certain number of doctors. Cross-sectional data may limit the range of explanations for price changes; future studies should address this limitation by extending the panel data to accurately simulate price changes over time while reducing time-sensitivity issues. Second, all of the empirical data were collected from the Haodf website. Although the choice in this study to use only one OHC helped us to improve internal validity, it can also reduce the generalizability of our findings, and future research should validate our results in other OHCs. In particular, samples of remote medical services provided by traditional hospitals can be included. It is worth studying whether the provision of online health care services by traditional hospitals or through internet companies can bring more benefits to society as a whole. Finally, due to the availability of data and to ethical considerations, we did not include any patient-related information but instead used only the doctors' public information. In future studies, under the condition of having obtained patients' privacy permission, researchers could discuss the factors affecting online health care consulting prices from the perspective of patient needs.

Conclusions

This study employed data collected from the Haodf website, which includes information related to 16,008 doctors, 1682 different hospitals, and 30 provinces in China. The main findings can be summarized as follows. Our empirical results support the conclusion that five level-1 variables (the doctor's eWOM volume, eWOM valence, past sales, clinic title, and hospital level) have significantly positive effects on the online health care consulting price. In other words, the doctor's online and offline reputation are both meaningful signals that affect their consulting price. In addition, the level-2 variable (provincial wage level) also had a significantly positive impact on the consulting price. The beauty of the internet is that it breaks through geographical restrictions and ideally makes it possible for anyone to access online services due to the relatively low threshold of information transmission. Nevertheless, the wage level in the geographical area in which the doctor works affects the charges for different online consulting services. This suggests that the internet can indeed reduce the cost of delivering information, but it might not eliminate the barriers leading to differences in urban and rural consumption levels.

Acknowledgments

We thank the editor and anonymous referee for very helpful comments and suggestions that greatly improved the paper. Y-TH gratefully acknowledges financial support from the Ministry of Education Project of Humanities and Social Sciences (20YJCZH199).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Robustness tests for the determinants of online health care consulting prices.
[DOCX File , 21 KB - jmir_v23i7e29170_app1.docx]

References

1. Danaher TS, Gallan AS. Service research in health care. J Service Res 2016 Sep 19;19(4):433-437 [FREE Full text] [doi: 10.1177/1094670516666346]
2. Gao GG, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: how do online ratings reflect population perceptions of quality? MIS Q 2015 Mar 3;39(3):565-589. [doi: 10.25300/misq/2015/39.3.03]
3. Telehealth Index: 2017 Consumer survey. American Well. URL: http://go.americanwell.com/rs/335-QLG-882/images/American_Well_Telehealth_Index_2017_Consumer_Survey.pdf [accessed 2021-06-20]
4. Dullet NW, Geraghty EM, Kaufman T, Kissee JL, King J, Dharmar M, et al. Impact of a University-based outpatient telemedicine program on time savings, travel costs, and environmental pollutants. Value Health 2017 Apr;20(4):542-546 [FREE Full text] [doi: 10.1016/j.jval.2017.01.014] [Medline: 28407995]
5. Health. OECD data. URL: https://data.oecd.org/health.htm [accessed 2021-06-20]
6. Goh JM, Gao G, Agarwal R. The creation of social value: can an online health community reduce rural-urban health disparities? MIS Q 2016 Jan 1;40(1):247-263. [doi: 10.25300/MISQ/2016/40.1.11]
7. Wang JN, Chiu YL, Yu H, Hsu YT. Understanding a nonlinear causal relationship between rewards and physicians' contributions in online health care communities: longitudinal study. J Med Internet Res 2017 Dec 21;19(12):e427 [FREE Full text] [doi: 10.2196/jmir.9082] [Medline: 29269344]
8. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
9. Wang L, Wu T, Guo X, Zhang X, Li Y, Wang W. Exploring mHealth monitoring service acceptance from a service characteristics perspective. Electron Commerce Res Appl 2018 Jul;30:159-168. [doi: 10.1016/j.elerap.2018.06.003]
10. The 39th china statistics report on internet development. China Internet Network Information Center (CNNIC). URL: http://www.cac.gov.cn/2017-01/22/c_1120352022.htm [accessed 2021-06-20]

XSL•FO
RenderX

11. Vargo SL, Lusch RF. The four service marketing myths. J Service Res 2016 Jun 29;6(4):324-335. [doi: 10.1177/1094670503262946]
12. Hoffman K, Turley L, Kelley SW. Pricing retail services. J Bus Res 2002 Dec;55(12):1015-1023. [doi: 10.1016/s0148-2963(00)00227-7]
13. Avlonitis GJ, Indounas KA. Pricing objectives and pricing methods in the services sector. J Serv Market 2005 Jan;19(1):47-57. [doi: 10.1108/08876040510579398]
14. McLachlin R. Service quality in consulting: what is engagement success? Manag Service Qual 2000 Jun;10(3):141-150. [doi: 10.1108/09604520010336669]
15. Momparler A, Carmona P, Lassala C. Quality of consulting services and consulting fees. J Bus Res 2015 Jul;68(7):1458-1462. [doi: 10.1016/j.jbusres.2015.01.033]
16. Lassala C, Carmona P, Momparler A. Alternative paths to high consulting fees: A fuzzy-set analysis. J Bus Res 2016 Apr;69(4):1367-1371. [doi: 10.1016/j.jbusres.2015.10.109]
17. Dulleck U, Kerschbamer R. On doctors, mechanics, and computer specialists: the economics of credence goods. J Econ Lit 2006 Feb 01;44(1):5-42. [doi: 10.1257/002205106776162717]
18. Dulleck U, Kerschbamer R, Sutter M. The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition. Am Econ Rev 2011 Apr 01;101(2):526-555. [doi: 10.1257/aer.101.2.526]
19. Pemer F, Skjølsvik T. The cues that matter: Screening for quality signals in the ex ante phase of buying professional services. J Bus Res 2019 May;98:352-365. [doi: 10.1016/j.jbusres.2019.02.005]
20. Holmstrom B. Moral hazard and observability. Bell J Econ 1979;10(1):74-91. [doi: 10.2307/3003320]
21. Pauly MV. Overinsurance and public provision of insurance: the roles of moral hazard and adverse selection. Quart J Econ 1974 Feb;88(1):44. [doi: 10.2307/1881793]
22. Doney PM, Cannon JP. An examination of the nature of trust in buyer–seller relationships. J Market 2018 Nov 19;61(2):35-51. [doi: 10.1177/002224299706100203]
23. Ba S, Pavlou PA. Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Q 2002 Sep;26(3):243. [doi: 10.2307/4132332]
24. Cheema A. Surcharges and seller reputation. J Consum Res 2008 Jun 01;35(1):167-177. [doi: 10.1086/529532]
25. Hofmann DA. An overview of the logic and rationale of hierarchical linear models. J Manag 2016 Jun 30;23(6):723-744. [doi: 10.1177/014920639702300602]
26. Raudenbush S, Bryk A. Hierarchical linear models: applications and data analysis methods. USA: Sage; 2002.
27. Schlissel MR, Chasin J. Pricing of services: an interdisciplinary review. Service Indust J 1991 Jul;11(3):271-286. [doi: 10.1080/02642069100000046]
28. Tung W, Capella L, Tat P. Service pricing: a multi-step synthetic approach. J Serv Mark 1997;11(1):53-65. [doi: 10.1080/08876049710795685]
29. Zeithaml VA, Parasuraman A, Berry LL. Problems and strategies in services marketing. J Market 1985;49(2):33-46. [doi: 10.2307/1251563]
30. Dearden J. Cost accounting comes to service industries. Harv Bus Rev 1978;56(5):132-140. [Medline: 10237527]
31. Arnold DR, Hoffman KD, McCormick J. Service pricing: a differentiation premium approach. J Serv Mark 1989 Mar;3(3):25-33. [doi: 10.1108/eum0000000002490]
32. Bonnici JL. Pricing dimensions in health care services. J Prof Serv Market 1992 Feb 12;8(1):57-66. [doi: 10.1300/j090v08n01_05]
33. Hoffman KD, Arnold D. Professional services pricing: An extended cost-oriented approach. Services Marketing Quart 1989;5(1):29-39. [doi: 10.1080/15332969.1989.9984936]
34. Anderson EW, Fornell C, Lehmann DR. Customer satisfaction, market share, and profitability: findings from Sweden. J Market 2018 Nov 28;58(3):53-66. [doi: 10.1177/002224299405800304]
35. Jiang Z, Benbasat I. Research Note—Investigating the influence of the functional mechanisms of online product presentations. Inf Syst Res 2007 Dec;18(4):454-470. [doi: 10.1287/isre.1070.0124]
36. Williamson OE. Calculativeness, trust, and economic organization. J Law Econ 1993 Apr;36(1, Part 2):453-486. [doi: 10.1086/467284]
37. Ganesan S. Determinants of long-term orientation in buyer-seller relationships. J Market 2018 Nov 28;58(2):1-19. [doi: 10.1177/002224299405800201]
38. Utz S, Matzat U, Snijders C. On-line reputation systems: the effects of feedback comments and reactions on building and rebuilding trust in on-line auctions. Int J Electron Comm 2014 Dec 08;13(3):95-118. [doi: 10.2753/jec1086-4415130304]
39. Hennig-Thurau T, Walsh G, Walsh G. Electronic word-of-mouth: motives for and consequences of reading customer articulations on the internet. Int J Electron Comm 2014 Dec 08;8(2):51-74. [doi: 10.1080/10864415.2003.11044293]
40. Duan W, Gu B, Whinston A. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. J Retail 2008 Jun;84(2):233-242. [doi: 10.1016/j.jretai.2008.04.005]
41. Liu Y. Word of mouth for movies: its dynamics and impact on box office revenue. J Market 2006 Jul;70(3):74-89. [doi: 10.1509/jmkg.70.3.74]

XSL•FO
RenderX

42. Chevalier JA, Mayzlin D. The effect of word of mouth on sales: online book reviews. J Market Res 2018 Oct 10;43(3):345-354. [doi: 10.1509/jmkr.43.3.345]
43. Cui G, Lui H, Guo X. The effect of online consumer reviews on new product sales. Int J Electron Comm 2014 Dec 08;17(1):39-58. [doi: 10.2753/jec1086-4415170102]
44. Goes PB, Lin M, Au Yeung C. "Popularity effect" in user-generated content: evidence from online product reviews. Inf Syst Res 2014 Jun;25(2):222-238. [doi: 10.1287/isre.2013.0512]
45. Shapiro C. Premiums for high quality products as returns to reputations. Quart J Econ 1983 Nov;98(4):659. [doi: 10.2307/1881782]
46. Bolton GE, Katok E, Ockenfels A. How effective are electronic reputation mechanisms? An experimental investigation. Manag Sci 2004 Nov;50(11):1587-1602. [doi: 10.1287/mnsc.1030.0199]
47. Clemons EK, Gao GG, Hitt LM. When online reviews meet hyperdifferentiation: a study of the craft beer industry. J Manag Inf Syst 2014 Dec 08;23(2):149-171. [doi: 10.2753/mis0742-1222230207]
48. Zhu F, Zhang X. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. J Market 2010 Mar 01;74(2):133-148. [doi: 10.1509/jm.74.2.133]
49. Banerjee AV. A simple model of herd behavior. Quart J Econ 1992 Aug 01;107(3):797-817. [doi: 10.2307/2118364]
50. Bikhchandani S, Hirshleifer D, Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades. J Polit Econ 1992 Oct;100(5):992-1026. [doi: 10.1086/261849]
51. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
52. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of China's Good Doctor website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
53. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
54. Wang Y, Wu H, Lei X, Shen J, Feng Z. The influence of doctors' online reputation on the sharing of outpatient experiences: empirical study. J Med Internet Res 2020 Dec 11;22(12):e16691 [FREE Full text] [doi: 10.2196/16691] [Medline: 33306028]
55. Mudambi SM, Schuff D. Research Note: What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Q 2010;34(1):185. [doi: 10.2307/20721420]
56. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manag Inf Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
57. Haodf. URL: http://www.haodf.com [accessed 2021-06-16]
58. Chen S, Guo X, Wu T, Ju X. Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. Inf Process Manag 2020 Sep;57(5):102253. [doi: 10.1016/j.ipm.2020.102253]
59. Yang H, Zhang X. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: panel data analysis. J Med Internet Res 2019 Mar 22;21(3):e12156 [FREE Full text] [doi: 10.2196/12156] [Medline: 30900997]
60. Wind Economic Database. URL: https://www.wind.com.cn/en/edb.html [accessed 2021-06-16]
61. Osborne J. Advantages of hierarchical linear modeling. Pract Assess 2000;7(1):1-3. [doi: 10.4135/9781506326139.n303]
62. James LR. Aggregation bias in estimates of perceptual agreement. J Appl Psychol 1982;67(2):219-229. [doi: 10.1037/0021-9010.67.2.219]
63. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. USA: Academic Press; 1988.
64. Hofmann DA, Gavin MB. Centering decisions in hierarchical linear models: implications for research in organizations. J Manag 2016 Jun 30;24(5):623-641. [doi: 10.1177/014920639802400504]
65. Liu S, Zhang M, Gao B, Jiang G. Physician voice characteristics and patient satisfaction in online health consultation. Inf Manag 2020 Jul;57(5):103233. [doi: 10.1016/j.im.2019.103233]

Abbreviations

eWOM: electronic word of mouth
HLM: hierarchical linear modeling
ICC: intraclass correlation coefficient
OHC: online healthcare community

XSL•FO
RenderX

Edited by G Eysenbach; submitted 29.03.21; peer-reviewed by X Gao, W Yowono; comments to author 11.04.21; revised version received 06.05.21; accepted 20.05.21; published 14.07.21.

Please cite as:
Chiu YL, Wang JN, Yu H, Hsu YT
Consultation Pricing of the Online Health Care Service in China: Hierarchical Linear Regression Approach
J Med Internet Res 2021;23(7):e29170
URL: https://www.jmir.org/2021/7/e29170
doi:10.2196/29170
PMID:34259643

©Ya-Ling Chiu, Jying-Nan Wang, Haiyan Yu, Yuan-Teng Hsu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Status Quo and Research Trends of Neurosurgical Departments in China: Bibliometric and Scientometric Analyses

Bowen Ni[1,2*], MD; Minyi He[2,3*], PhD; Bei Cao[2,4*], MD; Jianmin He[2,5], MD; Yawei Liu[1,2], MD; Zhen Zhao[2,5], MD

[1]Department of Neurosurgery, Nanfang Hospital, Southern Medical University, Guangzhou, China

[2]Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Southern Medical University, Guangzhou, China

[3]Clinical Medicine Education Center, Nanfang Hospital, Southern Medical University, Guangzhou, China

[4]Laboratories & Facilities Management Office, Southern Medical University, Guangzhou, China

[5]Institute of Scientific Research, Southern Medical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Zhen Zhao, MD
Institute of Scientific Research
Southern Medical University
1838 Guangzhou Dadao Bei Street
Guangzhou, 510515
China
Phone: 86 020 61648165
Email: nfykdxchengguo@126.com

Abstract

Background: Modern neurosurgery is a relatively young discipline characterized by finesse and complexity. In recent years, neurosurgery in China has made continuous developments, with long-term progress and outstanding discoveries in many aspects of the field.

Objective: This scientometric investigation aimed to comprehensively provide insight into the development trends of neurosurgery in China, to demonstrate how the field has evolved.

Methods: PubMed database was searched to retrieve relevant papers published between 1988 and 2018 from neurosurgery institutions in China. The database of the National Natural Science Foundation of China was also retrieved for funding information. Information (eg, year of publication, journal, institute of origin) and keywords were collected from each paper after removing duplicates and filtering unintentional words. Co-word analysis was performed on the papers' keywords, and a time distribution matrix of coexisting keywords in a given paper (ie, termed co-words) was established. Co-words were clustered according to their growth rate within years and visually presented with a mountain plot and a heatmap. Trends and potential subspecialties were identified, and each topic, represented either by a co-word from publications or funding from the National Natural Science Foundation of China during the period from 2011 to 2018, was collected and analyzed.

Results: Within 15,972 publications on neurosurgery from institutions in China, diagnostic image was found to coexist the most with other keywords. Cluster 0, represented by diagnostic image with retrospective study, contained emerging topics with great developmental potential and demonstrated high growth rates in recent years. This finding suggests that the topics represented in Cluster 0 may represent future areas of important neurosurgical research. We also found that the developmental trend of China's neurosurgical research is highly correlated with National Natural Science Foundation of China funding acquisition.

Conclusions: Co-word analysis and visualization results provided insight into the emerging research topics that are of vital importance, which can be used as a reference by neurosurgeons and researchers for future investigations. In this study, our analysis strategy based on co-word biclustering was able to clearly demonstrate current academic subject development; therefore, co-word biclustering is a reliable bibliometric analysis strategy.

(J Med Internet Res 2021;23(7):e25700) doi:10.2196/25700

KEYWORDS

neurosurgery; bibliometric analysis; co-word biclustering analysis; visualized analysis

XSL•FO
RenderX

Introduction

Modern neurosurgery, with a history of only 100 years, has been developing rapidly since its birth in the late 19th century. In past decades, modern neurosurgery has experienced three stages of development: classical neurosurgery, microneurosurgery, and minimally invasive neurosurgery [1]. Modern neurosurgery can be divided into neurotrauma, pediatric neurosurgery, functional neurosurgery, cerebrovascular disease, skull base lesions, brain tumors, spinal neurosurgery, and other subdisciplines [2]. After 1949, modern neurosurgery in China sprouted, and though it started later than it did in other countries, it has shown amazing growth, with an abundance of published papers [3].

For decades, the most common way for researchers to become familiar with the state of the art and development trends in research has been to read published reviews. However, the large amount of literature that can be found for a single topic makes it increasingly difficult for researchers to quickly find the areas of research they are interested in. In recent years, bibliometric analysis has become increasingly popular. It uses bibliometric characteristics, such as country, institution, journal, or author, to measure the contribution of a research field and to predict, in detail, the research trends or hotspots in a certain field [4,5].

Co-word analysis of keywords in a certain field can be performed, not only to gather detailed information about subdisciplines, but also, to detect intrinsic associations between them [6]. If keywords coexist in a paper, it is likely that the topics that they represent are related, to some extent [7]. Moreover, if a large number of papers contain coexisting keywords (ie, termed *co-words*) during a period of time or if the number of published papers has increased over time, it can represent an emerging area of interest that is currently, or will be in the future, highly important [8].

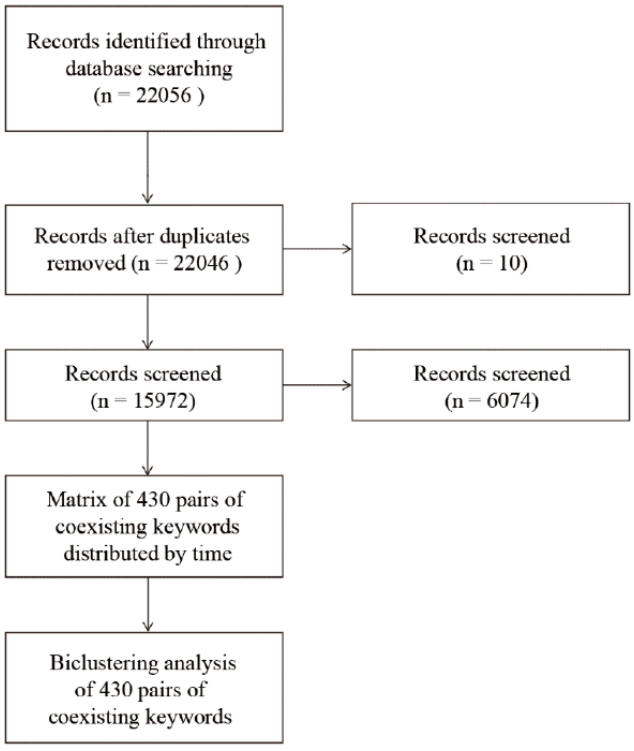
Cluster analysis, in particular, can help identify relationships between research topics. Bibliometric analysis has been used to identify the most influential papers in medical disciplines and specialties, such as gastric diseases [9], infectious diseases [10], and urinary diseases [11]; however, to date, literature trends from neurosurgery institutions in China have not been evaluated. We aimed to apply bibliometric and scientometric analysis techniques, such co-word matrixes and biclustering, to gain insight into domestic research on neurosurgery and subdisciplines, their evolution in recent years, and future development trends in China.

Methods

Search Methodology and Strategies

Published literature in the PubMed database was searched using the following search string: ((China [Affiliation]) AND neurosurgery [Affiliation]) AND (“1988” [Date - Publication]: “2018” [Date - Publication]). We performed the search on September 8, 2019, which yielded a total of 22,056 papers. Search results were screened to removed duplicates. We excluded articles about other subjects of study by title, abstract, and keywords (Figure 1). Article information, including the title, name of the journal, institution, year of publication, and keywords, was exported using the PubMed tool and saved as a document in .nbib format for subsequent analyses.

Figure 1. Data collection flowchart.



XSL•FO
RenderX

Data Processing

We imported bibliography data and keywords from the papers using NoteExpress (version 3.3; Beijing Aegean Software Company), and Excel (version 2019; Microsoft Inc) was used to establish a database containing bibliography information, such as the year of publication, influencing factors, and journals, and perform descriptive statistical analysis. The main premise of bibliometric analysis based on keywords is that keywords selected by paper authors fully and accurately describe the main content of the paper [8]. If a paper's keywords could not be extracted automatically by the software, we manually obtained the keywords from the paper in the PubMed database. If keywords were not available from the original source, we extracted subject headings from the title of the paper instead. The quality of keywords is of vital importance to the results of the subsequent analysis [12]. We standardized extracted keywords by (1) merging words with the same meaning, (2) removing duplicates, (3) changing plural to singular form, and (4) deleting unintentional words (such as numbers).

Co-Word Analysis

After processing, there was a total of 15,972 keywords. Because there were relatively few papers published before 2010, the period from 1988 to 2018 was divided into 11 stages (1988-2000, 2001-2005, 2006-2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, and 2018). Distribution features of all keywords for every stage were quantitatively assessed for co-word analysis. We used data-mining software (BibExcel [13]) to calculate the frequency of every co-word (pair of coexisting keywords appearing in each paper) to form a matrix [14]. To understand the status quo and developing trends of co-words over time, co-word matrixes for every period were imported into R software (version 3.6.1; R Foundation for Statistical Computing), and a binary temporal distribution matrix, with co-words as rows and stages of time as columns, was constructed and exported with a merging algorithm—the number in each cell represented the cumulative frequency of each pair in a given time stage. The growth rate of each co-word pair (change in the number between successive stages) was calculated to analyze development trends of every research topic represented by each pair, and 430 co-words with a sustained growth trend (growth rate over 0) in 2015 to 2016, 2016 to 2017, and 2017 to 2018 were identified.

Biclustering Analysis of Coexisting Words

gCLUTO software (version 1.0; Kerapis Lab) and OpenGL-based mountain visualization were used for biclustering analysis of co-words. We imported the binary matrix data of the 430 co-words into gCLUTO, which uses a clustering algorithm based on partitional clustering, to determine the best number of subgroups *k* based on height, color, and independence of each peak in the mountain plot.

The co-words were divided into 3 clusters, and co-word matrix data were shown by a peak plot and a heat map, which visually display distribution characteristics. Word cloud plots for the 3 clusters co-words were created, in which the size of each co-word was determined by its cumulative frequency. We also imported the co-word matrix into Cytoscape (version 3.8.2; National Institute of General Medical Sciences of the National Institutes of Health) to more intuitively reflect the relationships among 430 co-words with a circular layout.

Topic Trends Between Publications and Funding

The National Natural Science Foundation of China is the most important funding source in China; therefore, to provide insight into the role played by funding in development trends, funding amounts from 2011 to 2018 for each subdiscipline represented by co-words were collected, and Pearson correlations were calculated.

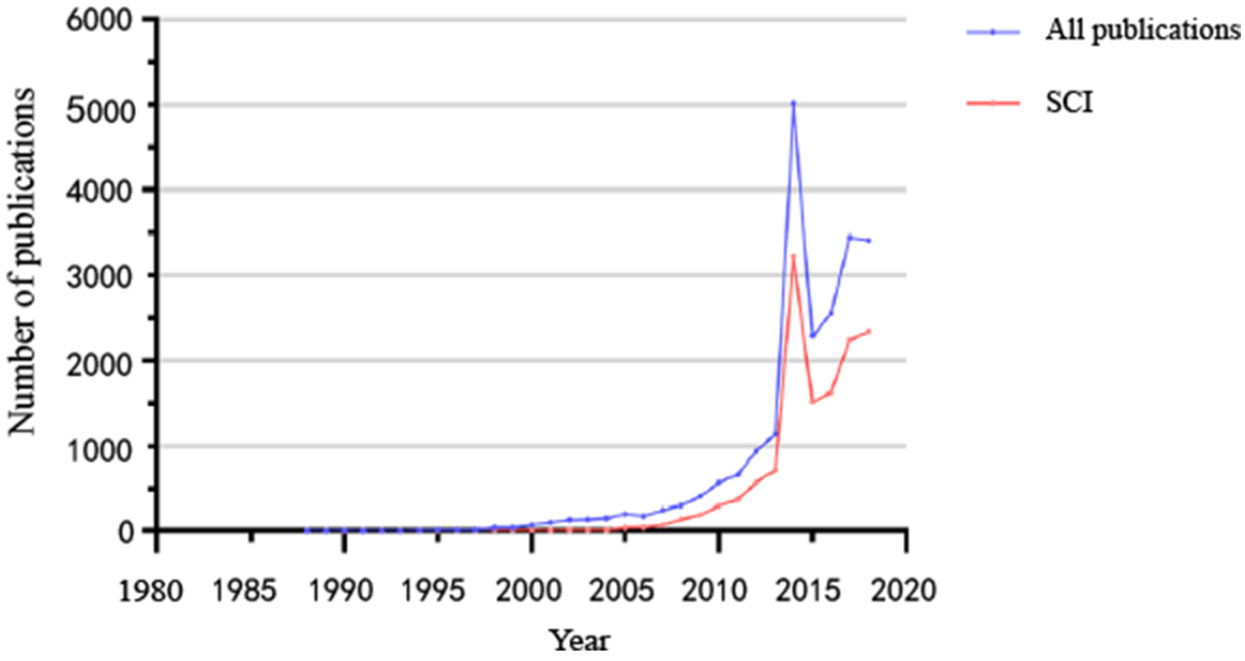
Results

Basic Characteristics of the Papers

Changes in the number of publications in a certain discipline can be examined, not only to reflect development, but also, to predict future developmental trends. To understand the development of neurosurgery in China in recent years, 15,972 papers from 1988 to 2018 were statistically analyzed year by year (Figure 2). The results showed that the amount of papers have increased year by year and reached a peak in 2014 (5021/15,972, 31.44%). The amount of papers included in the Science Citation Index (SCI), accounting for 61.00% (9743/15,972), has also been increasing rapidly, which is demonstrated by an upward trend since 2007.

XSL•FO
RenderX

Figure 2. The temporal distribution of articles. SCI: Science Citation Index.



Journal Distribution

Papers published by neurosurgical researchers in China appeared in 1420 journals, of which 802 are included in the SCI. The average number of papers published in the top 10 journals (Table 1) was 498, which was 32 times the number published in all other journals. *World Neurosurgery* featured the most papers (1096/15,972, 6.86%), followed by the *Chinese Medical Journal* (561/15,972, 3.51%) and *PLOS One* (481/15,972, 3.01%). Of the top 10 journals, 5 were included in the SCI, among which *PLOS One* had the highest impact factor (2.776). In total, 2906 papers were published in journals in the SCI (2906/4981, 58.34%), including those in the top 10 journals.

Table 1. Journals with the most published papers.

Rank	Journal	Papers, n
1	World Neurosurgery	1096
2	Zhonghua Yi Xue Za Zhi	561
3	PLOS One	481
4	Molecular Medicine Reports	478
5	Oncotarget	474
6	Journal of Clinical Neuroscience	444
7	Oncology Letters	409
8	Chinese Medical Journal (English)	382
9	Journal of Neuro-Oncology	329
10	Clinical Neurology and Neurosurgery	327

Institutions of Origin

Of scientific research institutions affiliated with the first author, 25.35% (4049/15,972) were from the top 10 institutions: Beijing Tiantan Hospital was the most prolific institution, with 985 papers (Table 2).

XSL•FO
RenderX

Table 2. Institutions with the most published papers.

Rank	Institution	Papers, n
1	Beijing Tiantan Hospital, Capital Medical University	985
2	Huashan Hospital, Fudan University	598
3	West China Hospital, Sichuan University	587
4	Jinling Hospital, Medical School of Nanjing University	388
5	Tangdu Hospital, Air Force Medical University	285
6	Peking Union Medical College Hospital	254
7	The General Hospital of Tianjin Medical University	248
8	Xijing Hospital, Fourth Military Medical University	240
9	Xiangya Hospital, Central South University	237
10	Changhai Hospital, Naval Medical University	227

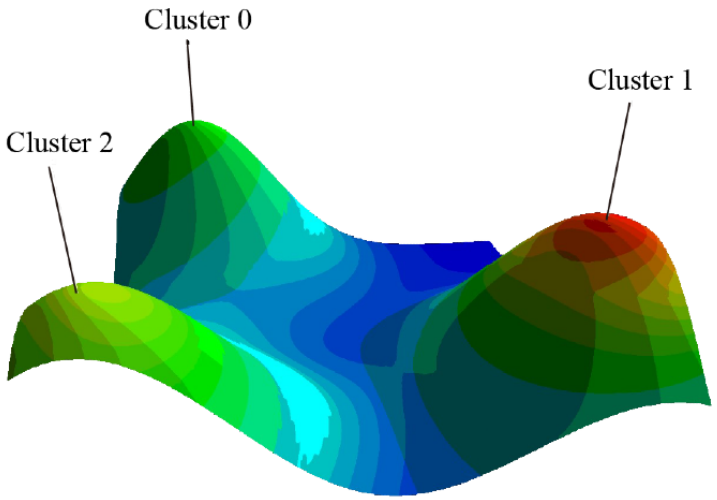
Co-word Analysis

A total of 24,469 keywords were extracted from the papers in the field of neurosurgery published by researchers in China from 1988 to 2018; 430 co-words had growth rates greater than 0 for 2015 to 2016, 2016 to 2017, and 2017 to 2018 (Multimedia Appendix 1). This filtering strategy was based on 430 pairs of co-words that can provide insight into emerging research subspecialties of great potential that have been attracting increasing attention within years and which might become the focus of neurosurgical research in the future.

Biclustering Analysis of 430 Pairs

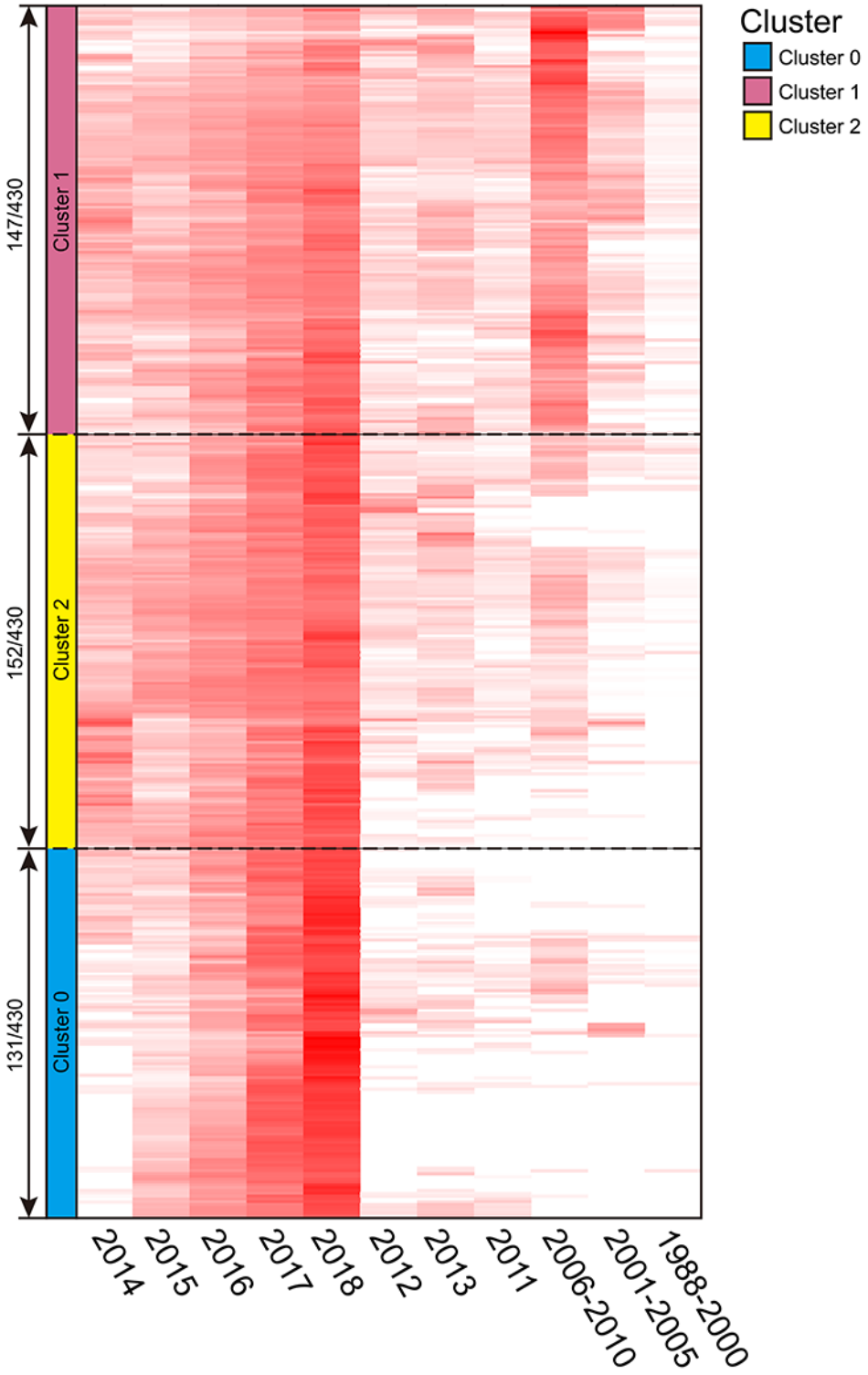
The clustering results achieved after many attempts to obtain the best clustering strategy are presented in the form of a peak plot (Figure 3), with 3 clusters marked as peaks 0 to 2, and a heatmap (Figure 4), which is a schematic representation of high-dimensional data [15]. The height, volume, and color of each peak is proportional to the internal similarity within each cluster, the number of co-words in the cluster, and the standard deviation among the co-words in one cluster, respectively [16,17]. The independent distribution of peaks indicates that the clustering strategy has performed well [18]. Interrelationship networks between 430 pairs of co-words showed that *diagnostic image* coexisted the most with other keywords (Figure 5).

Figure 3. Mountain visualization of biclustering of the 430 pairs of co-words. Red represents a low standard deviation, and blue represents a high standard deviation among the co-words in the cluster.



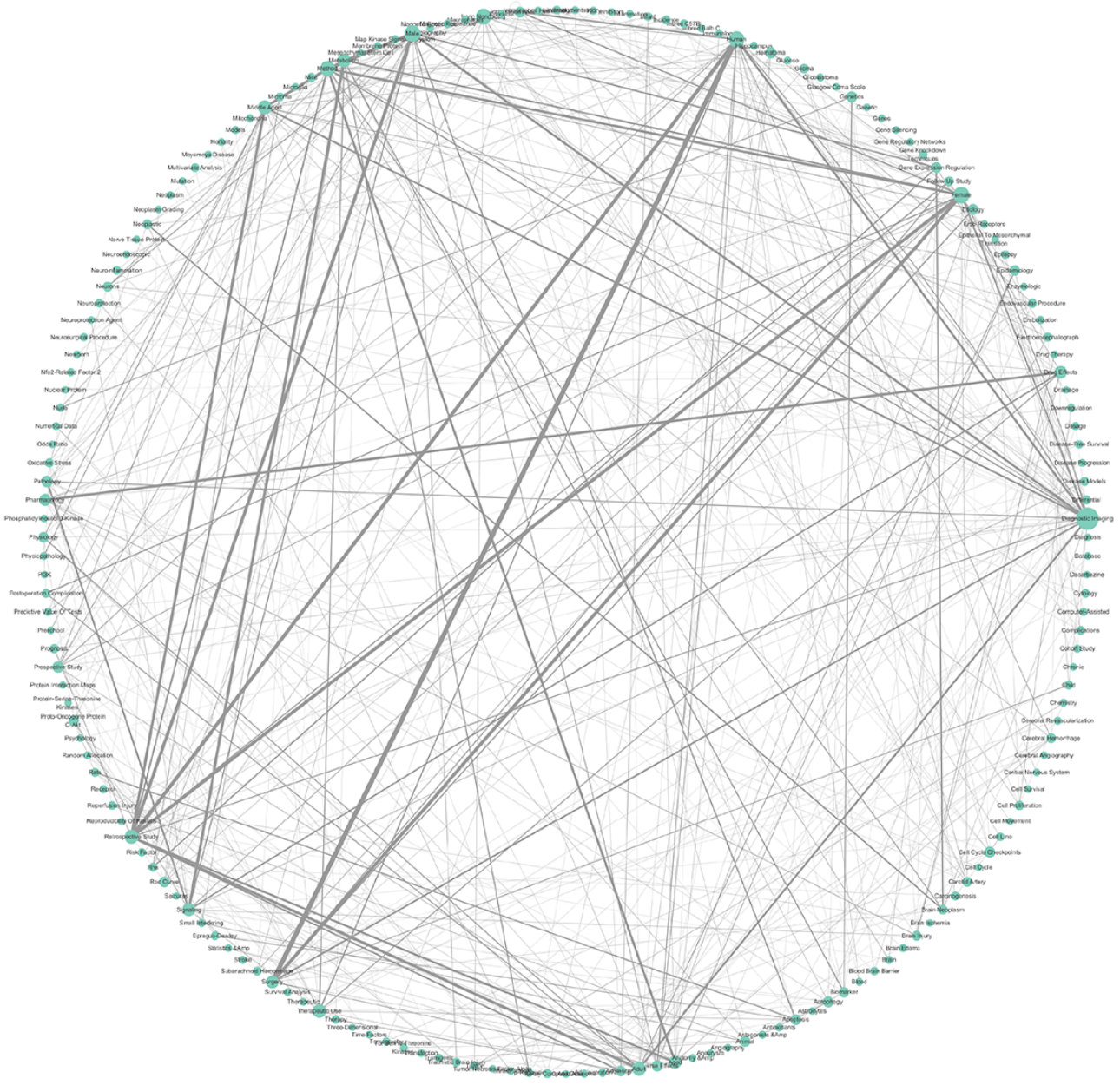
XSL•FO
RenderX

Figure 4. Heatmap of biclustering of the 430 pairs of co-words. Each row represents a pair of co-words, each column represents each stage of time, and the color of each box represents the frequency (the darker the red, the higher the frequency; white indicates that the frequency is close to zero [17]).



XSL•FO
RenderX

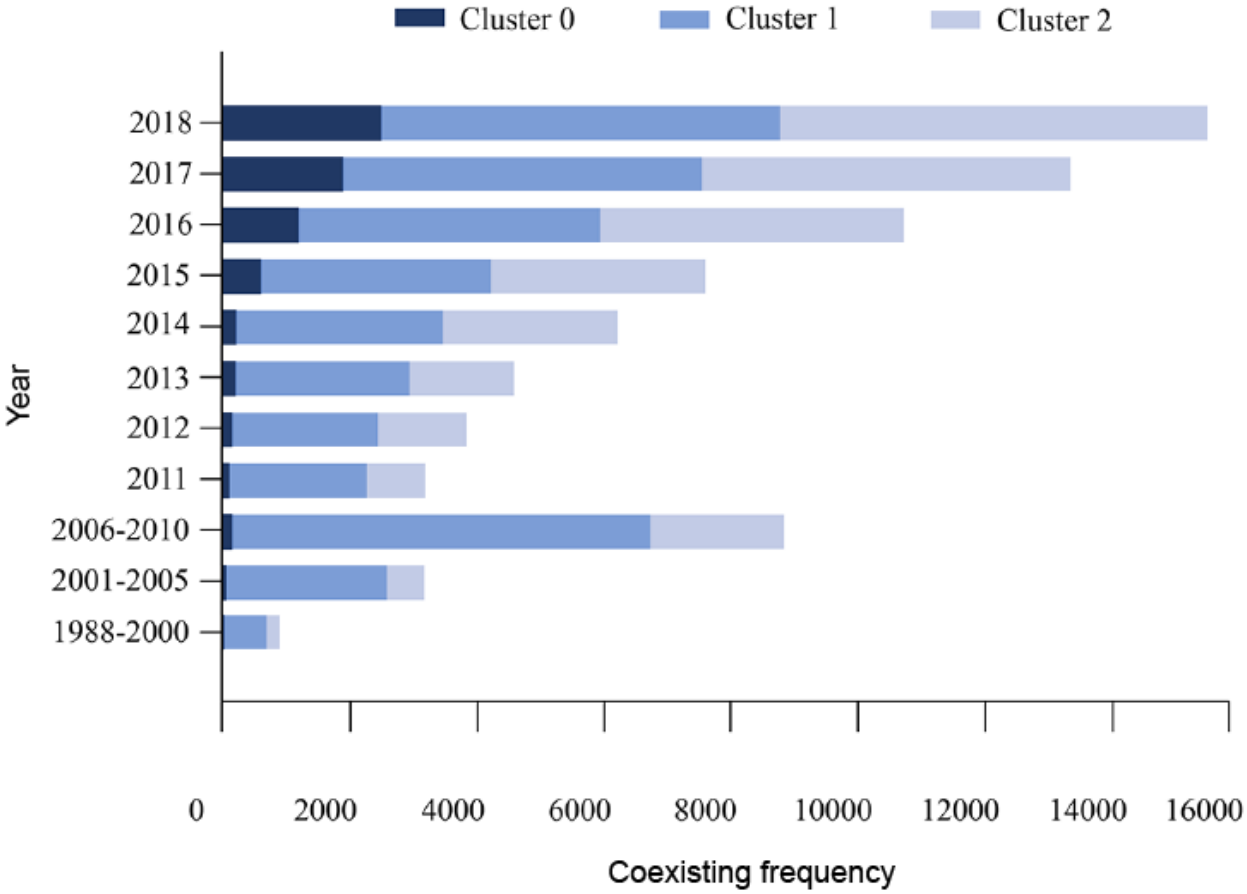
Figure 5. Relationships network among the 430 co-words. Each circle represents one keyword; the higher the frequency of coexistence, the larger the volume. Each line indicates a coexisting relationship between co-words (pairs of keywords); the higher the frequency of coexistence, the thicker the line.



The peak of Cluster 0 was the highest (internal similarity 0.945) has a large volume, which indicates that it contains a large number of co-words. The amount of papers containing the co-words in this cluster was low before 2015 (Figure 4 and Figure 6). Cluster 0 contained 131 co-words (total frequency: 7196/78,957, 9.11%). The most frequent co-words in Cluster 0 are (1) *diagnostic image* with *retrospective study* (428 papers), (2) *long noncoding* with *RNA* (208 papers), and (3) *long noncoding* with *human* (184 papers), respectively. These findings suggest that retrospective analyses of the imaging diagnosis of neurosurgical diseases and studies on human long noncoding RNA are leading areas in Cluster 0.

XSL•FO
RenderX

Figure 6. Temporal distribution of the co-word frequency for each cluster divided by stage.



The peak of Cluster 1 was the lowest (internal similarity 0.916) and had the largest volume, with 152 co-words. Cluster 1 contained a large number of papers containing these co-words in this cluster published since 2001 (total frequency: 41331/78957, 52.35%). The most frequent co-words in Cluster 1 were (1) *surgery* with *human* (2357 papers), (2) *human* with *retrospective study* (1752 papers), and (3) *surgery* with *female* (1615 papers). These results suggest that literature on neurosurgical treatment, retrospective analysis, and surgical treatment of neurosurgical diseases of female patients occupy the main positions in Cluster 1.

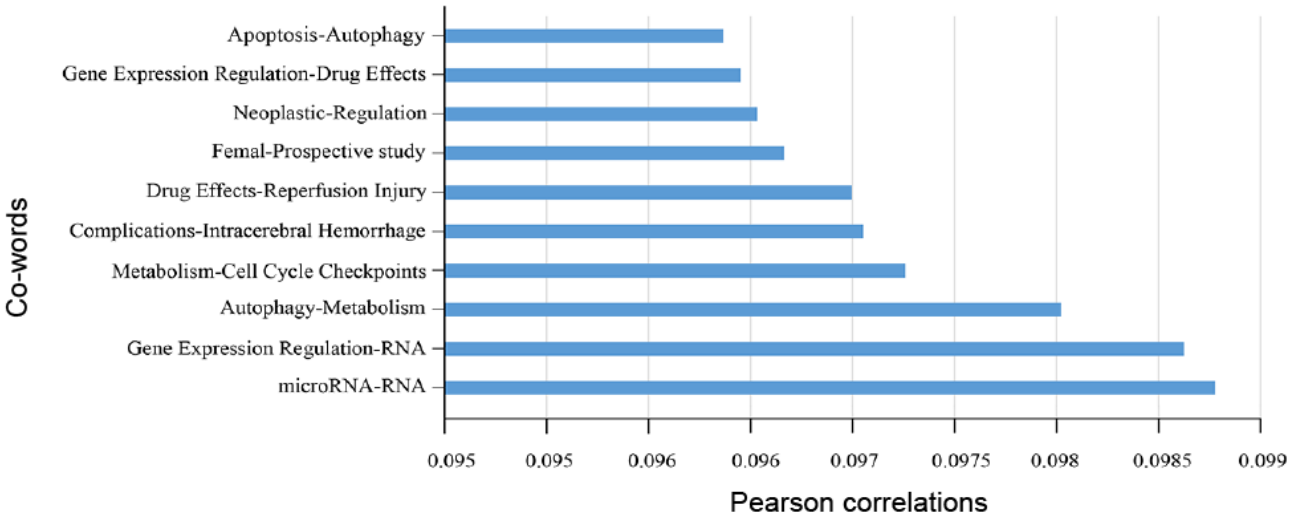
The peak of Cluster 2 was medium height (internal similarity 0.931), indicating great similarities among co-words in this category, and of medium volume, with 147 co-words (total frequency: 30,430/78,957, 38.54%). The number of papers in this cluster has grown rapidly since 2014 and continues to increase rapidly. The top most frequent co-words in this cluster were (1) *metabolism* with *signaling* (1425 papers), (2) *drug effects* with *pharmacology* (1225 papers), and (3) *diagnostic imaging* with *male* (1093 papers). We concluded that, in Cluster 2, literature related to metabolic signaling pathways, pharmacology, and papers about neurosurgical imaging diagnosis of male patients were ranked in the leading position.

Correlation Between the Number of Papers With Funding and Co-words

The distribution of National Natural Science Foundation of China funding for each topic of neurosurgical research represented by 430 co-words yearly (2011-2018) (Figure 7 and Multimedia Appendix 2). The topic *microRNA* was associated with the most papers funded by the National Natural Science Foundation of China (r=0.988), followed by *gene expression* with *RNA* (r=0.986) and *autophagy* with *metabolism* (r=0.980).

XSL•FO
RenderX

Figure 7. Correlation of the co-word frequency and number of papers with National Natural Science Foundation funds.



Discussion

Principal Findings

In this study, a bibliometric method, with keywords representing neurosurgery research papers from institutions in China published from 1988 to 2018 in the PubMed database, was used to analyze information on emerging topics. Quantitative analysis based on co-word and biclustering analyses of keywords in papers revealed development trends for each cluster of co-words and provides prospects for future research directions.

The bibliometric approach can help researchers address enormous amounts information when carrying out scientific investigations. The identification of potential research topics can promote academic advances in certain specialties and can help researchers grasp which emerging themes may become future directions of vital importance in certain disciplines [19].

In China, compared with the development of other fields, neurosurgery developed relatively late. Our study demonstrated a fluctuating and gradual increase in the amount of papers from China's neurosurgical institutions over time. The amount of publications showed a noticeable increase in recent years, with an explosive increment in 2014 that accounted for 23% (5021/15,972) of the total number of papers published from 1988 to 2018. While the number of publications showed an overall upward trend, the period from 1988 to 2006 exhibited steady growth, and the period from 2007 to 2018 exhibited rapid growth; the number of published papers in the latter period was 19 times higher than that in the former. The amount of papers from journals in the SCI increased over time, accounting for 61.00% (9743/15,972) of the total number of papers, which indicates development of the research and the influence of the field of neurosurgery in China.

The keywords of papers are representative [20]. To a large extent, keywords can provide insight into the main research contents of the literature accurately and briefly [21]. The accurate extraction of publication keywords is a very important technology in the field of scientific literature searching and data mining. At present, when extracting keywords from literature and establishing a keyword database, the manual method often obtains the best results [22]. We standardized and filtered the extracted keywords of papers and established a keyword database for papers published by neurosurgical institutions in China from 1988 to 2018 based on PubMed data.

The biclustering approach, in contrast to traditional clustering approaches, can simultaneously cluster year of publication and co-words into matrix columns and rows, respectively. Cluster 0, which contains topics that began to gradually develop in recent years, contained the least co-words. This finding reveals that topics represented by co-words included in Cluster 0 as emerging themes may grow in importance in the future, given that they appeared with increasing frequency in recent years. Retrospective analyses of diagnostic imaging data can, not only assist in clinical decision making, but also, improve the accuracy of neurosurgical disease diagnosis and predict response to disease treatment [23,24]. Long noncoding RNA plays an important role in the development of tumors. With advances in sequencing technology, many new types of long noncoding RNA, which cannot be translated into proteins, have been discovered in the nucleus and cytoplasm of human tumor cells, usually ranging from 200 nucleotides to 100 kilobases [25,26]. Long noncoding RNA can carry out many functions, such as cell growth, development, senescence, and death, by regulating gene expression. At present, an increasing number of studies have shown some long noncoding RNA are closely correlated with the existence and development of tumors [27].

Co-words contained in Cluster 1 and Cluster 2 exhibited large cumulative frequency counts, representing topics that have been studied much in the past, and the amounts have shown growth in recent years. As evidenced by Cluster 1, surgery is still the first choice for neurosurgery-related diseases, such as gliomas. However, traditional craniotomy is dependent on operator experience and visual judgment, and its application is limited because of the inevitable damage to normal tissue and blood vessels [28]. With the help of diagnostic imaging to accurately resect lesions, microscopic resection, which can significantly

XSL•FO
RenderX

improve clinical outcomes, has gradually become a main method of surgical treatment for human craniocerebral diseases [29,30]. As evidenced in Cluster 2, metabolic reprogramming of some craniocerebral tumors is closely related to signaling pathways, such as Wnt/β-catenin signaling pathways [31]. Metabolic changes in tumor cells are also intricately related to their characteristics, such as autonomous growth, antiapoptosis, unlimited proliferation, and angiogenesis [32].

Although the cumulative frequency counts of co-words in Cluster 0 were low, they increased from 2015 to 2018, which shows great potential for development; therefore, to a large extent, the topics in this cluster may become the future hotspots to which researchers should pay attention.

In general, funding facilitates scientific research. As the largest funding agency of natural science of China, the National Natural Science Foundation has made great efforts in encouraging the development of neurosurgical research over the past years. This study identified that the number of papers represented by the co-word pair *micro-RNA* and *RNA*, the co-word pair *gene expression regulation* and *RNA*, and the co-word pair *autophagy* and *metabolism* was proportional to the number of papers with National Natural Science Foundation in China funding, which may indicate that funding on these topics has accelerated research progress and literature output. Consistent co-word and biclustering analyses results suggest that research on the function of gene expression regulation of RNA, especially noncoding RNA such as micro-RNAs and long noncoding RNA, is of considerable importance and provide insight into future potential research directions.

Scientometric analysis can provide insight into the development of a research field, highlight research trends over time, and help identify valuable future directions. This study used bibliometric analysis to count and analyze the literature published by Chinese neurosurgical research institutions in the PubMed database from 1988 to 2018. Co-words were divided into 3 clusters that reveal current and future trends for neurosurgery research in China.

Limitations

This study has some limitations. First, we focused only on papers that were indexed in PubMed (a freely available database). Although we believe that PubMed is a sufficiently large database for publications from neurosurgical institutions in China, other databases, such as Web of Science, should be used in future research to obtain more information such as number of citations. Second, the search string was relatively general and may not be able to identify all papers on neurosurgery in China. In addition, because the number of studies has increased over time, the numbers of papers published in recent years may have had an effect on our results to some extent. Future investigations should be performed to provide a more comprehensive evaluation of papers on neurosurgery in China.

Conclusions

We applied bibliometric methods (co-word biclustering analysis and visual analysis) to examine research progression, emerging topics, and future potential research directions in the field of neurosurgery research in China. These findings can be used by neurosurgery researchers as a reference for choosing scientific research projects.

Acknowledgments

This study was supported by the Department of Education of Guangdong Province (2018WTSCX010).

Authors' Contributions

ZZ led the team and was responsible for all aspects of the project. BN and BC substantially contributed to the methods, data acquisition, results, and interpretation. MH participated in designing and writing the manuscript. JH revised this manuscript critically for important intellectual content. YL gave final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Temporal distribution of the 430 pairs of co-words.
[XLS File (Microsoft Excel File), 156 KB - jmir_v23i7e25700_app1.xls]

Multimedia Appendix 2
Temporal distribution of funding from the National Natural Scientific Foundation.
[XLS File (Microsoft Excel File), 82 KB - jmir_v23i7e25700_app2.xls]

References

1. Couldwell WT, Mortimer V, Kraus KL. Expanding neurosurgery. J Neurosurg 2014 Dec;121(6):1297-1313. [doi: 10.3171/2014.8.JNS141791] [Medline: 25434710]
2. Emejulu J, Malomo A, Oremakinde A, Onyia C, Nwaribe E, Ekweogwu O, et al. Developing a guideline for neurotrauma in Nigeria. World Neurosurg 2020 Jun;138:e705-e711. [doi: 10.1016/j.wneu.2020.03.036] [Medline: 32179184]

XSL•FO
RenderX

3. Li C, Zhu H, Zong X, Wang X, Gui S, Zhao P, et al. History, current situation, and future development of endoscopic neurosurgery in China. World Neurosurg 2018 Feb;110:270-275. [doi: 10.1016/j.wneu.2017.11.103] [Medline: 29180086]
4. Emery E. [Bibliometric analysis of neurosurgery publications in France]. Neurochirurgie 2019 Feb;65(1):7-13. [doi: 10.1016/j.neuchi.2018.08.006] [Medline: 30738601]
5. Alotaibi NM, Guha D, Fallah A, Aldakkan A, Nassiri F, Badhiwala JH, et al. Social media metrics and bibliometric profiles of neurosurgical departments and journals: is there a relationship? World Neurosurg 2016 Jun;90:574-579.e7. [doi: 10.1016/j.wneu.2016.01.087] [Medline: 26855310]
6. Li Y, He M, Zou Z, Bian X, Huang X, Yang C, et al. Artificial liver research output and citations from 2004 to 2017: a bibliometric analysis. PeerJ 2019;6:e6178 [FREE Full text] [doi: 10.7717/peerj.6178] [Medline: 30647995]
7. Kulakli A, Shubina I. Scientific publication patterns of mobile technologies and apps for posttraumatic stress disorder treatment: bibliometric co-word analysis. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e19391 [FREE Full text] [doi: 10.2196/19391] [Medline: 33242019]
8. Liu Y, Navathe SB, Pivoshenko A, Dasigi VG, Dingledine R, Ciliax BJ. Text analysis of MEDLINE for discovering functional relationships among genes: evaluation of keyword extraction weighting schemes. Int J Data Min Bioinform 2006;1(1):88-110. [doi: 10.1504/ijdmb.2006.009923] [Medline: 18402044]
9. Yuan F, Cai J, Liu B, Tang X. Bibliometric analysis of 100 top-cited articles in gastric disease. Biomed Res Int 2020 May 15;2020:1-8. [doi: 10.1155/2020/2672373]
10. Yang W, Zhang J, Ma R. The prediction of infectious diseases: a bibliometric analysis. Int J Environ Res Public Health 2020 Aug 27;17(17):6218 [FREE Full text] [doi: 10.3390/ijerph17176218] [Medline: 32867133]
11. Mainwaring A, Bullock N, Ellul T, Hughes O, Featherstone J. The top 100 most cited manuscripts in bladder cancer: a bibliometric analysis (review article). Int J Surg 2020 Mar;75:130-138. [doi: 10.1016/j.ijsu.2020.01.128] [Medline: 31991242]
12. Shen L, Wang S, Dai W, Zhang Z. Detecting the interdisciplinary nature and topic hotspots of robotics in surgery: social network analysis and Bibliometric Study. J Med Internet Res 2019 Mar 26;21(3):e12625 [FREE Full text] [doi: 10.2196/12625] [Medline: 30912752]
13. How to use BibExcel for various types of bibliometric analysis. BibExcel. URL: https://homepage.univie.ac.at/juan.gorraiz/bibexcel/ [accessed 2009-08-25]
14. Yang Y. Using bibliometric analysis to explore research trend of electronic word-of-mouth from 1999 to 2011. Int J Innov Technol 2012;3(4):337. [doi: 10.7763/ijimt.2012.v3.250]
15. Xu S, Shu G, Qiang S, Lei C, Xiumei Z. Advances in plastic and cosmetic surgery at home and abroad - a bibliometric analysis. Eur Rev Med Pharmacol Sci 2013 Oct;17(20):2732-2754 [FREE Full text] [Medline: 24174355]
16. Zhu X, Niu X, Li T, Liu C, Chen L, Tan G. Identification of research trends concerning application of stent implantation in the treatment of pancreatic diseases by quantitative and biclustering analysis: a bibliometric analysis. PeerJ 2019;7:e7674 [FREE Full text] [doi: 10.7717/peerj.7674] [Medline: 31660258]
17. Wei W, Shi B, Guan X, Ma J, Wang Y, Liu J. Mapping theme trends and knowledge structures for human neural stem cells: a quantitative and co-word biclustering analysis for the 2013-2018 period. Neural Regen Res 2019 Oct;14(10):1823-1832 [FREE Full text] [doi: 10.4103/1673-5374.257535] [Medline: 31169201]
18. Lu K, Yu S, Yu M, Sun D, Huang Z, Xing H, et al. Bibliometric analysis of tumor immunotherapy studies. Med Sci Monit 2018 May 23;24:3405-3414 [FREE Full text] [doi: 10.12659/MSM.910724] [Medline: 29790485]
19. Devos P, Menard J. Bibliometric analysis of research relating to hypertension reported over the period 1997-2016. J Hypertens 2019 Nov;37(11):2116-2122 [FREE Full text] [doi: 10.1097/HJH.0000000000002143] [Medline: 31136459]
20. Ouyang W, Wang Y, Lin C, He M, Hao F, Liu H, et al. Heavy metal loss from agricultural watershed to aquatic system: a scientometrics review. Sci Total Environ 2018 Oct 01;637-638:208-220. [doi: 10.1016/j.scitotenv.2018.04.434] [Medline: 29751304]
21. Shukla N, Merigó JM, Lammers T, Miranda L. Half a century of computer methods and programs in biomedicine: a bibliometric analysis from 1970 to 2017. Comput Methods Programs Biomed 2020 Jan;183:105075. [doi: 10.1016/j.cmpb.2019.105075] [Medline: 31526946]
22. Zhang J, Xie J, Hou W, Tu X, Xu J, Song F, et al. Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based co-word analysis and social network analysis. PLoS One 2012;7(4):e34497 [FREE Full text] [doi: 10.1371/journal.pone.0034497] [Medline: 22496819]
23. Chi C, Du Y, Ye J, Kou D, Qiu J, Wang J, et al. Intraoperative imaging-guided cancer surgery: from current fluorescence molecular imaging methods to future multi-modality imaging technology. Theranostics 2014;4(11):1072-1084 [FREE Full text] [doi: 10.7150/thno.9899] [Medline: 25250092]
24. Doris CI. Diagnostic imaging at its centennial: the past, the present and the future. CMAJ 1995 Nov 01;153(9):1297-1300 [FREE Full text] [Medline: 7497392]
25. Ma Y, Zhang J, Wen L, Lin A. Membrane-lipid associated lncRNA: a new regulator in cancer signaling. Cancer Lett 2018 Apr 10;419:27-29. [doi: 10.1016/j.canlet.2018.01.008] [Medline: 29330108]
26. Robinson EK, Covarrubias S, Carpenter S. The how and why of lncRNA function: an innate immune perspective. Biochim Biophys Acta Gene Regul Mech 2020 Apr;1863(4):194419 [FREE Full text] [doi: 10.1016/j.bbagrm.2019.194419] [Medline: 31487549]

XSL•FO
RenderX

27. St Laurent G, Wahlestedt C, Kapranov P. The landscape of long noncoding RNA classification. Trends Genet 2015 May;31(5):239-251 [FREE Full text] [doi: 10.1016/j.tig.2015.03.007] [Medline: 25869999]
28. Shoffstall AJ, Paiz JE, Miller DM, Rial GM, Willis MT, Menendez DM, et al. Potential for thermal damage to the blood-brain barrier during craniotomy: implications for intracortical recording microelectrodes. J Neural Eng 2018 Jun;15(3):034001 [FREE Full text] [doi: 10.1088/1741-2552/aa9f32] [Medline: 29205169]
29. Abolfotoh M, Bi WL, Hong C, Almefty KK, Boskovitz A, Dunn IF, et al. The combined microscopic-endoscopic technique for radical resection of cerebellopontine angle tumors. J Neurosurg 2015 Nov;123(5):1301-1311. [doi: 10.3171/2014.10.JNS141465] [Medline: 25909571]
30. Singh H, Essayed WI, Cohen-Gadol A, Zada G, Schwartz TH. Resection of pituitary tumors: endoscopic versus microscopic. J Neurooncol 2016 Nov;130(2):309-317. [doi: 10.1007/s11060-016-2124-y] [Medline: 27161249]
31. He L, Zhou H, Zeng Z, Yao H, Jiang W, Qu H. Wnt/β-catenin signaling cascade: a promising target for glioma therapy. J Cell Physiol 2019 Mar;234(3):2217-2228. [doi: 10.1002/jcp.27186] [Medline: 30277583]
32. Agnihotri S, Zadeh G. Metabolic reprogramming in glioblastoma: the influence of cancer metabolism on epigenetics and unanswered questions. Neuro Oncol 2016 Feb;18(2):160-172 [FREE Full text] [doi: 10.1093/neuonc/nov125] [Medline: 26180081]

Abbreviations

SCI: Science Citation Index

Edited by R Kukafka; submitted 12.11.20; peer-reviewed by Q Chen, J He, A Mavragani; comments to author 15.03.21; revised version received 25.04.21; accepted 04.05.21; published 05.07.21.

Please cite as:
Ni B, He M, Cao B, He J, Liu Y, Zhao Z
Status Quo and Research Trends of Neurosurgical Departments in China: Bibliometric and Scientometric Analyses
J Med Internet Res 2021;23(7):e25700
URL: https://www.jmir.org/2021/7/e25700
doi:10.2196/25700
PMID:36260378

©Bowen Ni, Minyi He, Bei Cao, Jianmin He, Yawei Liu, Zhen Zhao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment

Armando J Rotondi[1,2,3], DPhil; Jonathan Grady[4], DPhil; Barbara H Hanusa[1], DPhil; Gretchen L Haas[1,5], DPhil; Michael R Spring[6], DPhil; Kaleab Z Abebe[7], DPhil; James Luther[1], MA; John Gurklis[8], MD

[1]Mental Illness Research, Education and Clinical Center (MIRECC), VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

[2]Center for Health Equity Research and Promotion (CHERP), VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

[3]University of Pittsburgh Center for Behavioral Health, Media, and Technology, University of Pittsburgh, Pittsburgh, PA, United States

[4]Computer Science, Thomas College, Waterville, ME, United States

[5]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[6]Department of Information Sciences and Technology, School of Information Sciences, University of Pittsburgh, Pittsburgh, PA, United States

[7]Center for Research on Health Care Data Center, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[8]Behavioral Health, VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

Corresponding Author:
Armando J Rotondi, DPhil
Mental Illness Research, Education and Clinical Center (MIRECC)
VA Pittsburgh Health Care System
Department of Veterans Affairs
Research Office Bld, (151R-U)
University Drive C
Pittsburgh, PA, 15240
United States
Phone: 1 412 360 2369
Email: armandorotondi1@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/3/e23137

(J Med Internet Res 2021;23(7):e31553) doi:10.2196/31553

In “Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment” (J Med Internet Res 2021;23(3):e23137), two display errors were noted.

- In Table 1, row “Screen length,” column “Levels of dimensions in the factorial design; High,” the character “>” was not rendered in the originally published version. The value “1” has been corrected to “>1”.
- In Table 1, row “Constant navigational toolbar,” column “Levels of dimensions in the factorial design; High,” the value was not rendered in the originally published version. This value has been corrected to “1”.

In addition, five corrections have been made to the wording in Table 6:

- In row "The best performing websites had these design elements," "Mental illness other than SSD;" column "Design elements;" the phrase "screen per page length ≤1" has been corrected to "page length ≤1 screen long."
- In row "The best performing websites had these design elements," "No mental illness;" column "Design elements;" the phrase "screen per page length ≤1" has been corrected to "page length ≤1 screen long."
- In row "The worst performing websites with low navigational depth (≤3 levels) had these design elements," "Mental illness other than SSD;" column "Design elements;" the phrase "Screen per page length >1" has been corrected to "Page length >1 screen long."
- In row "Variables that, when present, always had a positive effect on performance," the subrow "Screen per page length ≤1" has been corrected to "Page length ≤1 screen long."
- In row "Variables that, when present, always had a negative effect on performance," the subrow "Screen length >1 screen long" has been corrected to "Page length >1 screen long."

XSL•FO
RenderX

The correction will appear in the online version of the paper on the JMIR Publications website on July 7, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.06.21; this is a non–peer-reviewed article;accepted 30.06.21; published 07.07.21.

Please cite as:
Rotondi AJ, Grady J, Hanusa BH, Haas GL, Spring MR, Abebe KZ, Luther J, Gurklis J
Correction: Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment
J Med Internet Res 2021;23(7):e31553
URL: https://www.jmir.org/2021/7/e31553
doi:10.2196/31553
PMID:34255689

©Armando J Rotondi, Jonathan Grady, Barbara H Hanusa, Gretchen L Haas, Michael R Spring, Kaleab Z Abebe, James Luther, John Gurklis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial

Richard Morriss[1], MD; Catherine Kaylor-Hughes[2], PhD; Matthew Rawsthorne[1], BA; Neil Coulson[3], PhD; Sandra Simpson[4], BSc; Boliang Guo[1], PhD; Marilyn James[3], PhD; James Lathe[5], MSc; Paul Moran[6], PhD; Laila J Tata[3], PhD; Laura Williams[1], PhD

[1]Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom
[2]Department of General Practice, University of Melbourne, Melbourne, United Kingdom
[3]School of Medicine, University of Nottingham, Nottingham, United Kingdom
[4]Research Delivery Team, Nottinghamshire Healthcare NHS Foundation Trust, Nottingham, United Kingdom
[5]Centre for Longitudinal Studies, University College London, London, United Kingdom
[6]School of Medicine, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Richard Morriss, MD
Institute of Mental Health
University of Nottingham
Jubilee Campus
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 1158230427
Email: richard.morriss@nottingham.ac.uk

Related Article:

Correction of: http://jmir.org/2021/4/e23487

(J Med Internet Res 2021;23(7):e31543) doi:10.2196/31543

In "A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial" (J Med Internet Res 2021;23(4):e23487), one error was noted.

Due to a system error, the name of one author, Laila J Tata, was replaced with the name of another author on the paper, Laura Williams. In the originally published paper, the order of authors was listed as follows:

> *Richard Morriss, Catherine Kaylor-Hughes, Matthew Rawsthorne, Neil Coulson, Sandra Simpson, Boliang Guo, Marilyn James, James Lathe, Paul Moran, Laura Williams, Laura Williams*

This has been corrected to:

> *Richard Morriss, Catherine Kaylor-Hughes, Matthew Rawsthorne, Neil Coulson, Sandra Simpson, Boliang Guo, Marilyn James, James Lathe, Paul Moran, Laila J Tata, Laura Williams*

In the originally published paper, the ORCID of author Laura Williams was incorrectly published as follows:

> *0000-0002-6404-8658*

This has been corrected to:

> *0000-0001-6296-6677*

The correction will appear in the online version of the paper on the JMIR Publications website on July 12, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 24.06.21; this is a non–peer-reviewed article;accepted 28.06.21; published 13.07.21.

Please cite as:
Morriss R, Kaylor-Hughes C, Rawsthorne M, Coulson N, Simpson S, Guo B, James M, Lathe J, Moran P, Tata LJ, Williams L
Correction: A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial
J Med Internet Res 2021;23(7):e31543
URL: https://www.jmir.org/2021/7/e31543
doi:10.2196/31543
PMID:34255667

©Richard Morriss, Catherine Kaylor-Hughes, Matthew Rawsthorne, Neil Coulson, Sandra Simpson, Boliang Guo, Marilyn James, James Lathe, Paul Moran, Laila J Tata, Laura Williams. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource

Konrad Krawczyk[1], DPhil; Tadeusz Chelkowski[2], MSc; Daniel J Laydon[3], PhD; Swapnil Mishra[3], PhD; Denise Xifara[4], DPhil; Benjamin Gibert[4], MSc; Seth Flaxman[5], DPhil; Thomas Mellan[3], PhD; Veit Schwämmle[6], PhD; Richard Röttger[1], PhD; Johannes T Hadsund[1], MSc; Samir Bhatt[3,7], DPhil

[1]Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

[2]Department of Management in the Network Society, Kozminski University, Warsaw, Poland

[3]Department of Infectious Disease Epidemiology, MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

[4]Nupinion, London, United Kingdom

[5]Department of Mathematics, Imperial College London, London, United Kingdom

[6]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

[7]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Konrad Krawczyk, DPhil
Department of Mathematics and Computer Science
University of Southern Denmark
Campusvej 55
Odense, DK-5230
Denmark
Phone: 45 6550 2387
Email: konradk@imada.sdu.dk

Related Article:

Correction of: https://www.jmir.org/2021/6/e28253

(J Med Internet Res 2021;23(7):e31544) doi:10.2196/31544

In "Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource" (J Med Internet Res 2021;23(6):e28253), one error was noted.

Due to a system error, the name of one author, Benjamin Gibert, was replaced with the name of another author on the paper, Seth Flaxman. In the originally published paper, the order of authors was listed as follows:

> *Konrad Krawczyk, Tadeusz Chelkowski, Daniel J Laydon, Swapnil Mishra, Denise Xifara, Seth Flaxman, Seth Flaxman, Thomas Mellan, Veit Schwämmle, Richard Röttger, Johannes T Hadsund, Samir Bhatt*

This has been corrected to:

> *Konrad Krawczyk, Tadeusz Chelkowski, Daniel J Laydon, Swapnil Mishra, Denise Xifara, Benjamin Gibert, Seth Flaxman, Thomas Mellan, Veit Schwämmle, Richard Röttger, Johannes T Hadsund, Samir Bhatt*

In the originally published paper, the ORCID of author Benjamin Gibert was incorrectly published as follows:

> *0000-0002-2477-4217*

This has been corrected to:

> *0000-0001-8457-3137*

The correction will appear in the online version of the paper on the JMIR Publications website on July 14, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 24.06.21; this is a non–peer-reviewed article;accepted 28.06.21; published 14.07.21.

Please cite as:
Krawczyk K, Chelkowski T, Laydon DJ, Mishra S, Xifara D, Gibert B, Flaxman S, Mellan T, Schwämmle V, Röttger R, Hadsund JT, Bhatt S
Correction: Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource
J Med Internet Res 2021;23(7):e31544
URL: https://www.jmir.org/2021/7/e31544
doi:10.2196/31544
PMID:34260375

©Konrad Krawczyk, Tadeusz Chelkowski, Daniel J Laydon, Swapnil Mishra, Denise Xifara, Benjamin Gibert, Seth Flaxman, Thomas Mellan, Veit Schwämmle, Richard Röttger, Johannes T Hadsund, Samir Bhatt. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Olfactory Training and Visual Stimulation Assisted by a Web Application for Patients With Persistent Olfactory Dysfunction After SARS-CoV-2 Infection: Observational Study

Fabrice Denis[1], MD, PhD; Anne-Lise Septans[2], PhD; Lea Periers[3], MD; Jean-Michel Maillard[4], ING; Florian Legoff[5], MSc; Hirac Gurden[6], PhD; Sylvain Moriniere[3], MD, PhD

[1]Institut Inter-Regional Jean Bernard - ELSAN, Le Mans, France

[2]Weprom, Angers, France

[3]Service d'Otorhinolaryngologie, Centre Hospitalier Universitaire Bretonneau, Tours, France

[4]anosmie.org Association, Alencon, France

[5]Kelindi, Lille, France

[6]Unite de Biologie Fonctionnelle Adaptative, Unite Mixte de Recherche 8251 Centre National de Recherche Scientifique, Université de Paris, Paris, France

Corresponding Author:
Fabrice Denis, MD, PhD
Institut Inter-Regional Jean Bernard - ELSAN
9 Rue Beauverger
Le Mans
France
Phone: 33 684190480
Email: fabrice.denis.home@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/5/e29583

(J Med Internet Res 2021;23(7):e32120) doi:10.2196/32120

In “Olfactory Training and Visual Stimulation Assisted by a Web Application for Patients With Persistent Olfactory Dysfunction After SARS-CoV-2 Infection: Observational Study” (J Med Internet Res 2021;23(5):e29583), one error was noted.

In the originally published article, the “Edited by” credit was listed incorrectly. This has been corrected from “Edited by C Basch, G Eysenbach” to “Edited by G Eysenbach”.

The correction will appear in the online version of the paper on the JMIR Publications website on July 16, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.07.21; this is a non–peer-reviewed article;accepted 15.07.21; published 16.07.21.

Please cite as:
Denis F, Septans AL, Periers L, Maillard JM, Legoff F, Gurden H, Moriniere S
Correction: Olfactory Training and Visual Stimulation Assisted by a Web Application for Patients With Persistent Olfactory Dysfunction After SARS-CoV-2 Infection: Observational Study
J Med Internet Res 2021;23(7):e32120
URL: https://www.jmir.org/2021/7/e32120
doi:10.2196/32120
PMID:34270442

XSL•FO
RenderX

©Fabrice Denis, Anne-Lise Septans, Lea Periers, Jean-Michel Maillard, Florian Legoff, Hirac Gurden, Sylvain Moriniere. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study

Cheng-Sheng Yu[1,2,3,4*], PhD; Shy-Shin Chang[1,2*], MD, PhD; Tzu-Hao Chang[3,5], PhD; Jenny L Wu[1,3], BSc; Yu-Jiun Lin[1,2], MD; Hsiung-Fei Chien[6], MD, PhD; Ray-Jade Chen[4,7,8], MD, MSc

[1]Department of Family Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[2]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[3]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[4]Professional Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[5]Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan

[6]Division of Plastic Surgery, Department of Surgery, Taipei Medical University Hospital and School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[7]Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[8]Division of General Surgery, Department of Surgery, Taipei Medical University Hospital, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Ray-Jade Chen, MD, MSc
Department of Surgery
School of Medicine, College of Medicine
Taipei Medical University
No.250, Wuxing St.,
Taipei, 11031
Taiwan
Phone: 886 227372181 ext 3966
Email: rayjchen@tmu.edu.tw

Related Article:

Correction of: https://www.jmir.org/2021/5/e27806

(J Med Internet Res 2021;23(7):e31085) doi:10.2196/31085

In “A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study” (J Med Internet Res 2021;23(5):e27806), the authors noticed two errors.

In the originally published manuscript, authors Cheng-Sheng Yu and Shy-Shin Chang should have been denoted as having contributed equally to the paper but were not. A footnote clarifying equal contribution has now been added to each of the aforementioned authors.

Additionally, in the originally published manuscript, the Acknowledgments section was omitted. In the corrected version, a new Acknowledgments section has been added with the following statement:

This study is supported by the Ministry of Science and Technology Grant (MOST 110-2314-B-038-025) and Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (DP2-110-21121-01-A-09). No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the article.

The correction will appear in the online version of the paper on the JMIR Publications website on July 9, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

XSL•FO
RenderX

Submitted 09.06.21; this is a non–peer-reviewed article;accepted 10.06.21; published 09.07.21.

Please cite as:
Yu CS, Chang SS, Chang TH, Wu JL, Lin YJ, Chien HF, Chen RJ
Correction: A COVID-19 Pandemic Artificial Intelligence–Based System With Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study
J Med Internet Res 2021;23(7):e31085
URL: https://www.jmir.org/2021/7/e31085
doi:10.2196/31085
PMID:34255678

©Cheng-Sheng Yu, Shy-Shin Chang, Tzu-Hao Chang, Jenny L Wu, Yu-Jiun Lin, Hsiung-Fei Chien, Ray-Jade Chen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study

Lizong Deng[1,2*], PhD; Luming Chen[1,2*], MS; Tao Yang[1,2*], MM; Mi Liu[3], PhD; Shicheng Li[1,2], MM; Taijiao Jiang[1,2,4], PhD

[1]Center of Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Suzhou Institute of Systems Medicine, Suzhou, China

[3]Jiangsu Institute of Clinical Immunology, Jiangsu Key Laboratory of Clinical Immunology, The First Affiliated Hospital of Soochow University, Suzhou, China

[4]Guangzhou Laboratory, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Taijiao Jiang, PhD
Center of Systems Medicine
Institute of Basic Medical Sciences
Chinese Academy of Medical Sciences & Peking Union Medical College
#5 Dong Dan San Tiao
Dongcheng District
Beijing, 100005
China
Phone: 86 051262873781
Email: taijiao@ibms.pumc.edu.cn

Related Article:

Correction of: https://www.jmir.org/2021/6/e26892

(J Med Internet Res 2021;23(7):e31481) doi:10.2196/31481

In “Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study” (J Med Internet Res 2021;23(6):e26892) the authors noted three errors.

1. In the originally published manuscript, affiliation 1, 2, and 3 were incorrectly mentioned as follows:

> *[1]Suzhou Institute of Systems Medicine, Suzhou, China*
>
> *[2]Jiangsu Institute of Clinical Immunology, Jiangsu Key Laboratory of Clinical Immunology, The First Affiliated Hospital of Soochow University, Suzhou, China*
>
> *[3]Center of Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*

These have been corrected to the following:

> *[1]Center of Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*
>
> *[2]Suzhou Institute of Systems Medicine, Suzhou, China*
>
> *[3]Jiangsu Institute of Clinical Immunology, Jiangsu Key Laboratory of Clinical Immunology, The First Affiliated Hospital of Soochow University, Suzhou, China*

2. The affiliation *"Center of Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China"* was originally listed only for author *Taijiao Jiang*. This affiliation has now been added to authors *Lizong Deng*, *Luming Chen*, *Tao Yang*, and *Shicheng Li* in addition to author *Taijiao Jiang*.

3. The address of the corresponding author was originally published as follows:

> *Chongwen Road 100*
>
> *Suzhou Industrial Park*
>
> *Suzhou, 215000*
>
> *China*

XSL•FO
RenderX

This has been corrected to the following:

#5 Dong Dan San Tiao
Dongcheng District
Beijing, 100005
China

The correction will appear in the online version of the paper on the JMIR Publications website on July 13, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.06.21; this is a non–peer-reviewed article;accepted 24.06.21; published 13.07.21.

Please cite as:
Deng L, Chen L, Yang T, Liu M, Li S, Jiang T
Correction: Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study
J Med Internet Res 2021;23(7):e31481
URL: https://www.jmir.org/2021/7/e31481
doi:10.2196/31481
PMID:34255682

©Lizong Deng, Luming Chen, Tao Yang, Mi Liu, Shicheng Li, Taijiao Jiang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology

Ji-Hyun Lee[1], RN, PhD; Hyeoun-Ae Park[2], RN, PhD; Tae-Min Song[3], PhD

[1]Nurse's Office, Yeongdeok High School, Gyeonggi-do, Republic of Korea

[2]College of Nursing and Research Institute of Nursing Science, Seoul National University, Seoul, Republic of Korea

[3]Department of Health Management, Samyook University, Seoul, Republic of Korea

Corresponding Author:
Hyeoun-Ae Park, RN, PhD
College of Nursing and Research Institute of Nursing Science
Seoul National University
103 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 27408827
Fax: 82 27661852
Email: hapark@snu.ac.kr

Related Article:

Correction of: https://www.jmir.org/2021/6/e25028

(J Med Internet Res 2021;23(7):e31601) doi:10.2196/31601

In "A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology" (J Med Internet Res 2021;23(6):e25028), the authors noted four errors.

Three errors were in the section "Applying the Ontology to Detect Future Signals" in the *Methods* section of the originally published article.

First, the originally published subheading:

> *Step 1. Collecting News*

has been corrected to:

> *Step 1. Collecting Data*

Second, curly brackets were missing from the following equation for DoD:



This has been corrected to:



Third, the descriptions of the first and second quadrants were incorrect in the following sentence:

> *Keywords in the first quadrant, which represent weak signals, have a low average DF but a high average DoD growth rate, and so they may increase rapidly in the future. Keywords in the second quadrant, which represent strong signals, have a trend toward a high average DF and a high average DoD growth rate.*

This has been corrected to:

> *Keywords in the first quadrant, which represent strong signals, have a trend toward a high average DF and a high average DoD growth rate. Keywords in the second quadrant, which represent weak signals, have a low average DF but a high average DoD growth rate, and so they may increase rapidly in the future.*

The fourth error was in the section "Applying the Ontology to Detect Future Signals" in the *Results* section of the originally published article.

In the original paper, the words "first" and "second" were switched in the following paragraph:

> *The weak signals are marked with red rectangle (area A) in the first quadrant of the KIM. The strong signals are marked with blue rectangle (area B) in the second quadrant of the KIM.*

This has been corrected to:

> *The weak signals are marked with red rectangle (area A) in the second quadrant of the KIM. The strong signals are marked with blue rectangle (area B) in the first quadrant of the KIM.*

XSL•FO
RenderX

The correction will appear in the online version of the paper on the JMIR Publications website on July 13, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 27.06.21; this is a non–peer-reviewed article;accepted 28.06.21; published 13.07.21.

Please cite as:
Lee JH, Park HA, Song TM
Correction: A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology
J Med Internet Res 2021;23(7):e31601
URL: https://www.jmir.org/2021/7/e31601
doi:10.2196/31601
PMID:34255676

©Ji-Hyun Lee, Hyeoun-Ae Park, Tae-Min Song. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predictive Modeling of Morbidity and Mortality in Patients Hospitalized With COVID-19 and its Clinical Implications: Algorithm Development and Interpretation

Joshua M Wang[1,2,3], BSc; Wenke Liu[1,2], PhD; Xiaoshan Chen[4], MD; Michael P McRae[5], PhD; John T McDevitt[5], PhD; David Fenyö[1,2,6], PhD

[1]Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY, United States
[2]Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY, United States
[3]Vilcek Institute of Graduate Biomedical Sciences, NYU Grossman School of Medicine, New York, NY, United States
[4]Department of Medicine, NYU Grossman School of Medicine, New York, NY, United States
[5]Department of Biomaterials, Bioengineering Institute, New York University, New York, NY, United States
[6]NYU Langone Health, New York, NY, United States

Corresponding Author:
David Fenyö, PhD
NYU Langone Health
435 E 30th St 9th Floor, Room 907
New York, NY, 10016
United States
Phone: 1 917 754 6066
Email: david@fenyolab.org

Abstract

Background: The COVID-19 pandemic began in early 2021 and placed significant strains on health care systems worldwide. There remains a compelling need to analyze factors that are predictive for patients at elevated risk of morbidity and mortality.

Objective: The goal of this retrospective study of patients who tested positive with COVID-19 and were treated at NYU (New York University) Langone Health was to identify clinical markers predictive of disease severity in order to assist in clinical decision triage and to provide additional biological insights into disease progression.

Methods: The clinical activity of 3740 patients at NYU Langone Hospital was obtained between January and August 2020; patient data were deidentified. Models were trained on clinical data during different parts of their hospital stay to predict three clinical outcomes: deceased, ventilated, or admitted to the intensive care unit (ICU).

Results: The XGBoost (eXtreme Gradient Boosting) model that was trained on clinical data from the final 24 hours excelled at predicting mortality (area under the curve [AUC]=0.92; specificity=86%; and sensitivity=85%). Respiration rate was the most important feature, followed by SpO_2 (peripheral oxygen saturation) and being aged 75 years and over. Performance of this model to predict the deceased outcome extended 5 days prior, with AUC=0.81, specificity=70%, and sensitivity=75%. When only using clinical data from the first 24 hours, AUCs of 0.79, 0.80, and 0.77 were obtained for deceased, ventilated, or ICU-admitted outcomes, respectively. Although respiration rate and SpO_2 levels offered the highest feature importance, other canonical markers, including diabetic history, age, and temperature, offered minimal gain. When lab values were incorporated, prediction of mortality benefited the most from blood urea nitrogen and lactate dehydrogenase (LDH). Features that were predictive of morbidity included LDH, calcium, glucose, and C-reactive protein.

Conclusions: Together, this work summarizes efforts to systematically examine the importance of a wide range of features across different endpoint outcomes and at different hospitalization time points.

(J Med Internet Res 2021;23(7):e29514) doi:10.2196/29514

KEYWORDS

COVID-19; coronavirus; SARS-CoV-2; predictive modeling; New York City; prediction; model; machine learning; morbidity; mortality; hospital; marker; severity; symptom; decision making; outcome

XSL•FO
RenderX

Introduction

The first cluster of SARS-CoV-2 cases was reported in Wuhan, Hubei Province, China, on December 31, 2019. With symptoms remarkably similar to pneumonia, the disease quickly traveled around the world, earning its pandemic status by the World Health Organization on March 11, 2020. Although the first wave has since passed for the hardest-hit regions, such as New York City and most of Asia, a resurgence of cases has already been reported in Europe and a record number of new cases has been tallied in the Midwest and rural United States. As of November 12, 2020, the United States alone logged its highest tally to date, with a 317% growth over the preceding 30 days [1]. COVID-19 is far from seeing the end of its days and there remains a compelling need to prioritize care and resources for patients at elevated risk of morbidity and mortality.

Previous work building machine learning models used patient data from Tongji Hospital in Wuhan, China [2,3]; Zhongnan Hospital in Wuhan, China [4]; Mount Sinai Hospital in New York City, United States [5]; and NYU (New York University) Family Health Center in New York City, United States [6]. Surprisingly, clinical features that were selected varied widely across studies. For example, while McRae et al's two-tiered model [6] that was trained on 701 patients in New York City to predict mortality was based on actual age, C-reactive protein (CRP), procalcitonin, and D-dimer, Yan et al's model [2] that was trained on 485 patients from Wuhan selected lactate dehydrogenase (LDH), lymphocyte count, and CRP as the most predictive for mortality. Variations in selected features differed greatly, even when trained to predict similar outcomes on data from patients of the same city. Yao et al's model [3] was trained on 137 patients from Wuhan, and the final model relied on 28 biomarkers to predict morbidity. Given the differences among prior models, some of which were driven by domain-specific knowledge, we decided to systematically examine the importance of a wide range of features across different endpoint outcomes and at different hospitalization time points.

This study analyzed retrospective polymerase chain reaction (PCR)–confirmed data from inpatients with COVID-19 that were collected at NYU Langone Hospital, spanning from January 1 to August 7, 2020, to predict three sets of clinical outcomes: alive versus deceased, ventilated versus not ventilated, or intensive care unit (ICU) admitted versus not ICU admitted. The clinical information of 3740 patient encounters included demographic data (ie, age, sex, insurance, past diagnosis of diabetes, and presence of cardiovascular comorbidities), vital signs (ie, SpO_2 [peripheral oxygen saturation], pulse, respiration rate, temperature, systolic blood pressure, and diastolic blood pressure), and the 50 most frequently ordered lab tests in our data set. Models were developed using two methods: logistic regression with feature selection using the least absolute shrinkage and selection operator (LASSO) [7] and gradient tree boosting with XGBoost (eXtreme Gradient Boosting) [8]. An explainable algorithm, such as logistic regression, provides easy-to-interpret insights into the features of importance. Conversely, the larger model capacity of XGBoost better handles data complexities to explore the extent to which predictive performance can be optimized. Together, this study aimed to provide a holistic survey of the clinical underpinnings of disease etiology for patients with COVID-19 admitted to NYU Langone Hospital. In addition, we sought to explore the prospects of building models that are sufficiently competent to be effective decision support tools.

Methods

Ethics Statement

An ethics exemption and a waiver were confirmed through the Institutional Review Board (IRB) at NYU Grossman School of Medicine. An IRB self-certification form was completed to ensure that the subsequent research did not fall under human subject research; therefore, no IRB approval was required. The deidentified COVID-19 NYU Langone Database was stripped of all unique identifiers prior to receiving data. In addition, all dates were shifted by an arbitrary number of days for each patient. These safeguards ensured that patient data could not be reidentified; thus, they were not subject to Health Insurance Portability and Accountability Act (HIPAA) restrictions on research use and did not require IRB approval.

Data Collection

The clinical activity of patients at NYU Langone Hospital was obtained from Epic—electronic medical record system—between January 1 and August 7, 2020. The data were stripped of all unique identifiers (medical record numbers, names, etc) and actual dates were shifted by an arbitrary number of days for each patient, which ensured that no data were subject to HIPAA restrictions and, thus, did not require IRB approval.

Clinical Data Preprocessing and Cleaning

Overview

Our data set contained 206,677 patients who were tested for COVID-19, of which 12,473 (6.0%) tested positive (Multimedia Appendix 1). Not all patients who tested positive sought hospital care, and without vital signs or lab values, these patients were excluded from analysis. In addition, a majority of the 175,507 patients diagnosed with COVID-19 did not receive in-house PCR tests, which makes it difficult to distinguish which hospital encounters were related to seeking COVID-19 treatment. Thus, only patients for which we could confirm a positive PCR test as reported by NYU Langone Hospital were included. The time stamp of the first encounter in which a PCR test returned a positive result was used as the starting date for each patient, and the ending date was determined by either the time of discharge for that encounter or the time of death. The clinical features that were collected for each patient, along with their definitions and additional processing steps, are described in the following subsections.

Categorical Features

The categorical features collected for each patient are listed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Categorical features.

Binned ages—to comply with Health Insurance Portability and Accountability Act restrictions on research use, exact patient ages were removed and binned into predefined ranges, as determined by the Data Handling Committee:

- 0-17 years
- 18-44 years
- 45-64 years
- 65-74 years
- 75+ years

Gender:

- 0 for female
- 1 for male

Insurance type:

- 0 for preferred provider organization
- 1 for exclusive provider organization, health maintenance organization, point-of-service-plan, indemnity, Medicare, Medicare managed care, no fault, and workers' compensation
- 2 for Medicaid and Medicaid managed care

Diabetes:

- 1 for any past diagnosis mentioning diabetes
- 0 otherwise

Cardiovascular comorbidities:

- 1 for any of the following ICD-10 (International Statistical Classification of Diseases and Related Health Problems, 10th Revision) diagnosis codes: I10-I16 (hypertensive diseases), I20-I25 (ischemic heart diseases), I50 (heart failure), I60-I69 (cerebrovascular diseases), and I72 (other aneurysms)
- 0 otherwise

Continuous Features

For each of the following continuous features (Textbox 2), multiple periodic measurements were recorded for each patient by vital signs monitors. Measurements were binned into 24-hour windows that began from time of hospitalization. Within each window, values were averaged. Values were then standardized to a mean of 0 and variance of 1. For each day, encounters without all features listed in Textbox 2 were removed and were not imputed.

Textbox 2. Continuous features.

- SpO_2 (peripheral oxygen saturation) (%)
- Pulse (bpm [beats per minute])
- Respiration rate (bpm)
- Temperature (°F)
- Systolic blood pressure (mm Hg)
- Diastolic blood pressure (mm Hg)

Outcomes

The outcomes for each patient are listed in Textbox 3.

XSL•FO
RenderX

Textbox 3. Patient outcomes.

Living status:

- 0 for alive
- 1 for dead

Ventilation at any point during hospitalization:

- 0 for no (did not receive any form of ventilation or only received noninvasive treatments; eg, nasal cannula, nonrebreather mask, etc)
- 1 for yes (received mechanical ventilation treatment)

Intensive care unit admission for any duration during hospitalization—criteria determined by medical triage team, balanced between disease severity and hospital resource availability:

- 0 for no
- 1 for yes

Lab Data Selection and Cleaning

Lab tests with at least 50% completeness during the first 24 hours for all encounters were considered. Of the 54 lab tests meeting these requirements, the estimated glomerular filtration rate—non-African and African American—was removed due to the formula's dependency on lab features already selected (ie, creatinine). In addition, the placeholders for ordering a complete blood count with differential test and a COVID-19 PCR test were also removed. Missing lab values were imputed using the multivariate imputation by chained equations algorithm. Five imputations were generated using predictive mean matching. After imputation, lab values were shifted up by 1 and log-transformed. Model-building approaches that incorporated lab features had individual models built for each imputation.

Feature Selection and Model Building

All models were trained with a training data to validation data ratio split of 90:10. Features for logistic regression were selected using LASSO and optimized for a penalty parameter that was 1 standard error above the minimum deviance for additional shrinkage. The XGBoost parameters were identified using a hyper-parameter search within the following constraints: nrounds=1000; η=0.3, 0.1, or 0.01; max_depth=2, 3, 4, 5, 6, 7, or 8; min_child_weight=0 to 1, by 0.1 increments; and γ=0 to 1, by 0.1 increments. To account for class imbalance, sample-weighted loss was employed when calculating the loss.

For models that were trained on the final day of discharge or death, the performance on predicting outcomes in all preceding days was evaluated on the entire data set rather than just a 10% subset. Data from previous days were not used in the training of these endpoint models and, thus, can all serve as validation data.

Time Series Modeling

In each feature setting, all variables were combined and missing values at each time point were imputed with the immediate previous value (ie, forward filling). After imputation, time points with incomplete feature measurements were discarded, and each patient record was segmented into nonoverlapping sequences of length 8. Patients were randomly assigned to training, validation, and testing groups in an 8:1:1 ratio for three independent splits. All models were implemented in Python 3.6 (Python Software Foundation) with built-in units in TensorFlow 2 and Keras [9]. Logistic regression was fit as a neural network with the sigmoid output node immediately after the input layer. For multilayer perceptron (MLP), recurrent neural network (RNN), gated recurrent unit (GRU), and long short-term memory (LSTM) models, a hidden layer of size 8 was added, and the time series models (ie, RNN, GRU, and LSTM) were unrolled over eight time points and trained with true labels provided at each step. Five randomly initialized models were trained for all architectures on each training, validation, and testing split. Model performance was evaluated based on all single time point predictions and reported as a mean value across all splits.

Data Availability

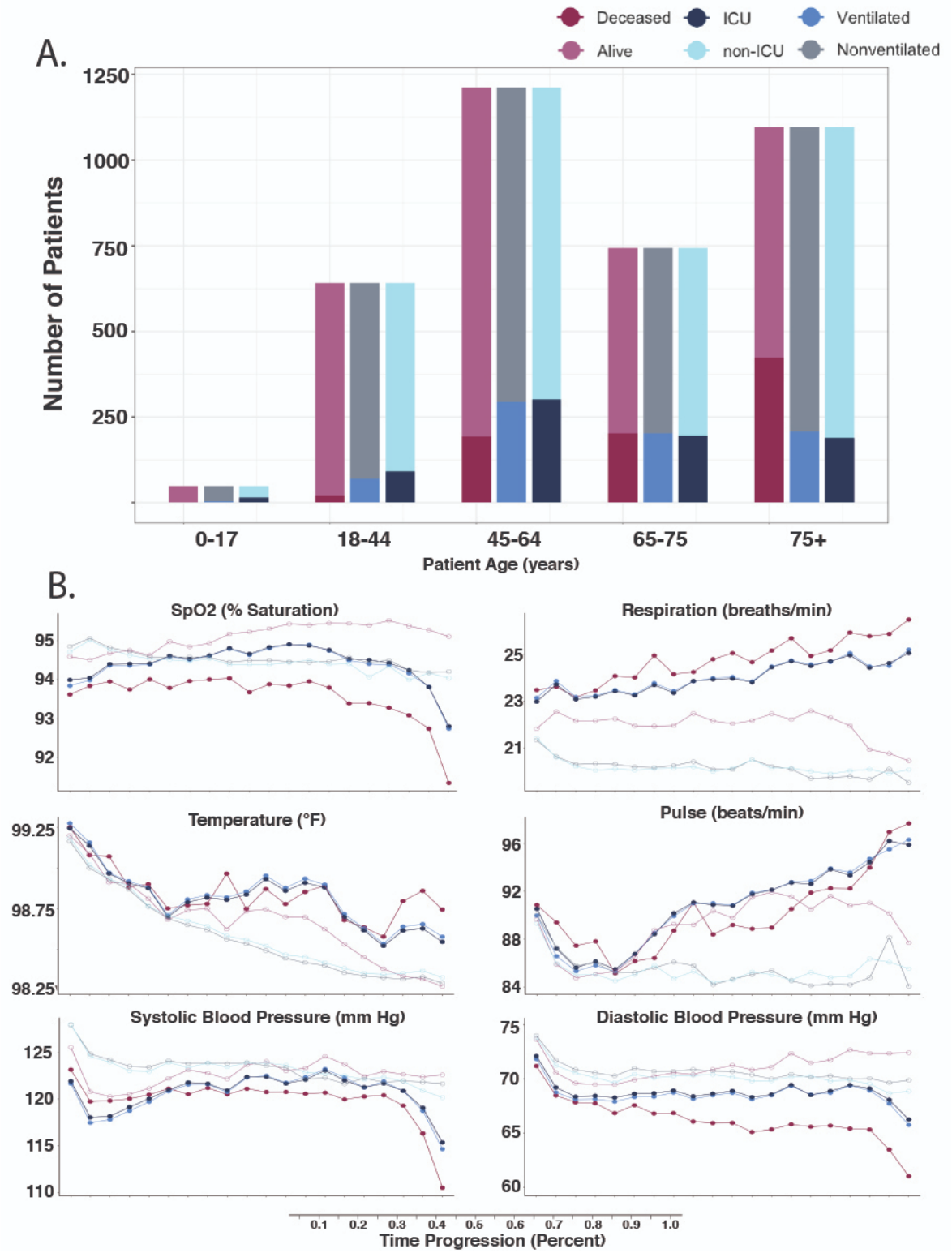
The data that support the findings of this study were obtained from the Medical Center Information Technology (MCIT) at NYU Langone Health, but restrictions apply to the availability of these data and, therefore, they are not publicly available due to specific institutional requirements.

Results

More than half of all patients in our data set were over the age of 65 years, with pediatric patients (0-17 years) having the lowest representation (Figure 1A and Multimedia Appendix 2). Generally, the proportion of deceased patients increased with age, peaking at 38.5% (422/1097) for those 75 years and over, 15.9% (193/1211) for those 45 to 64 years, and 0% for pediatric patients. Most patients who were either ventilated or admitted to the ICU belonged to the 65-to-74-years age group, followed by those 45 to 64 years and 75 years and over.

XSL•FO
RenderX

Figure 1. Overview of the clinical data set. A. Patient ages were binned by predefined ranges and the ratio of outcomes compared across age groups. B. For each patient, hospitalization stay was normalized by length of stay and segmented into 5% windows. Within each window, all values for the measured clinical variable were averaged. Each line is colored by the six possible outcomes. ICU: intensive care unit; SpO_2: peripheral oxygen saturation.



Aggregation of values for commonly acquired clinical metrics over normalized time courses offered meaningful insights into disease progression. Each patient's hospitalization stay was segmented into 5% windows, and clinical metrics were averaged within each bin (Figure 1B). We first examined the difference of average vital sign measurements between cohorts with different outcomes. The value of SpO_2 was statistically different for all three outcome comparisons in the first 5% of hospitalization time (Wilcoxon test: $P<.001$, $P<.001$, and $P<.001$). Over the clinical time course, the difference in SpO_2 averages increased the most for those who would become deceased, followed by those who were ICU admitted and ventilated. Differences in respiration rate followed a similar adverse trend, with breaths per minute increasing the most for

XSL•FO
RenderX

those who would become deceased, followed by those who were ventilated and ICU admitted. The divergence was present even after accounting for overlapping deceased patients. When considering the subset of patients who survived, ventilated patients had 2.91 more breaths per minute (Wilcoxon test: $P<.001$), and ICU-admitted patients had 2.90 more breaths per minute (Wilcoxon test: $P<.001$). At the beginning of the time course, differences in temperature were small (0.05 °F, 0.11 °F, and 0.06 °F, respectively) and not statistically significant for those who would become deceased (Wilcoxon test: $P=.13$), but differences were statistically significant for those who were ventilated (Wilcoxon test: $P<.001$) or were admitted to ICU (Wilcoxon test: $P<.001$). Pulse differences at the beginning were not significantly different for those who were ventilated (Wilcoxon test: $P=.29$) but were significantly different for those who would become deceased (Wilcoxon test: $P<.001$) and ICU admitted (Wilcoxon test: $P<.001$). Systolic and diastolic blood pressure values were continuously lower for patients with worse outcomes in this data set.

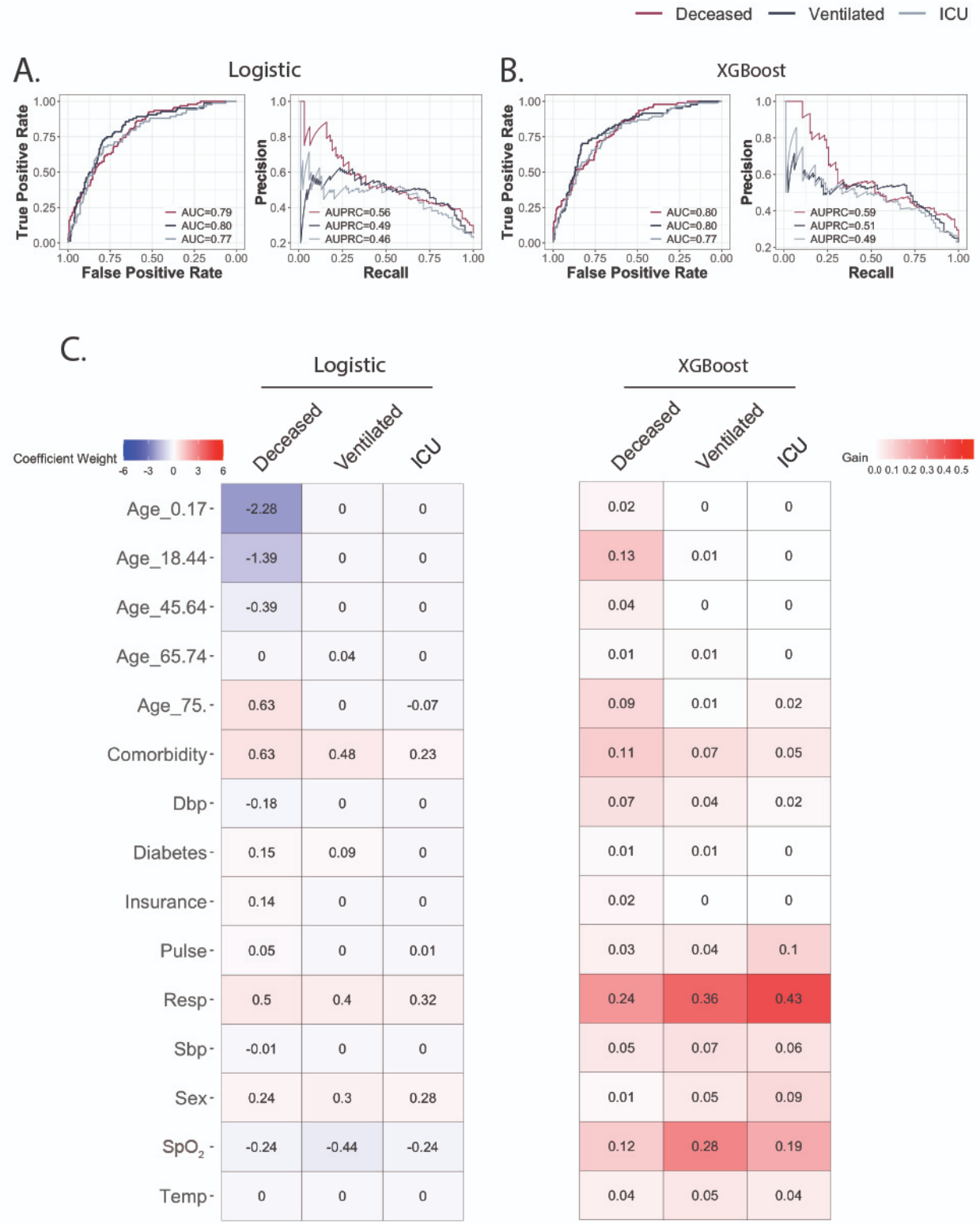
To assess the effectiveness of these vital signs to triage clinical outcomes, only data collected in the first 24 hours after admission were initially considered. Specifically for the ventilation outcome, respiration rates and SpO_2 levels may be influenced by the time point when mechanical ventilation was administered. Of 3740 encounters, 7.0% (262/3740) were ventilated within the first 24 hours of admission. To assess the bias that early administration of mechanical ventilation during the first 24 hours may have on respiration rate and SpO_2 levels, the distribution of values was compared against a filtered subset containing only values recorded prior to the start of ventilation. At the per-encounter level, the difference in respiration rates (Wilcoxon test: $P=.26$; Multimedia Appendix 3, plot A) and SpO_2 levels (Wilcoxon test: $P=.20$; Multimedia Appendix 3, plot B) were not significantly different. Given (1) that 93.0% (3478/3740) of encounters were not influenced by early ventilation treatment, (2) the insignificant difference in distributions, and (3) the desire to keep feature selection consistent across models, all values recorded within the first 24 hours were included. For each encounter, continuous features with multiple recordings (ie, SpO_2, pulse, respiration rate, temperature, systolic blood pressure, and diastolic blood pressure) were averaged and then standardized to a mean of 0 and a standard deviation of 1.

For logistic regression, features were selected using LASSO with 10-fold cross-validation. Grid search was used to optimize XGBoost parameters (Multimedia Appendix 4). When trained on data from the first 24 hours, the logistic model had area under the curve (AUC) performances of 0.79, 0.80, and 0.77; specificities of 59%, 78%, and 79%; and sensitivities of 86%, 74%, and 68%, respectively (Figure 2A). XGBoost performed similarly, with AUC performances of 0.80, 0.80, and 0.77; specificities of 59%, 83%, and 69%; and sensitivities of 86%, 70%, and 77%, respectively (Figure 2B).

In both logistic regression and gradient tree boosting settings, features of importance varied across clinical outcomes (Figure 2C). For logistic regression models of the three outcomes, respiration rate, SpO_2, and cardiovascular comorbidity were among predictive features, but age groups were selected only for predicting mortality. For boosting tree models, feature importance measures showed that respiration rate was consistently the most important feature for all three outcomes, and the age category 18 to 44 years was the second most important feature only for vital status. Respiration rate and SpO_2 were important for predicting all three outcomes. Differences in temperature were not strongly predictive in any cohort in either model; this finding and temperature's insignificant difference in the deceased outcome group together suggest that its role in screening for increased disease severity may not be dependable.

XSL•FO
RenderX

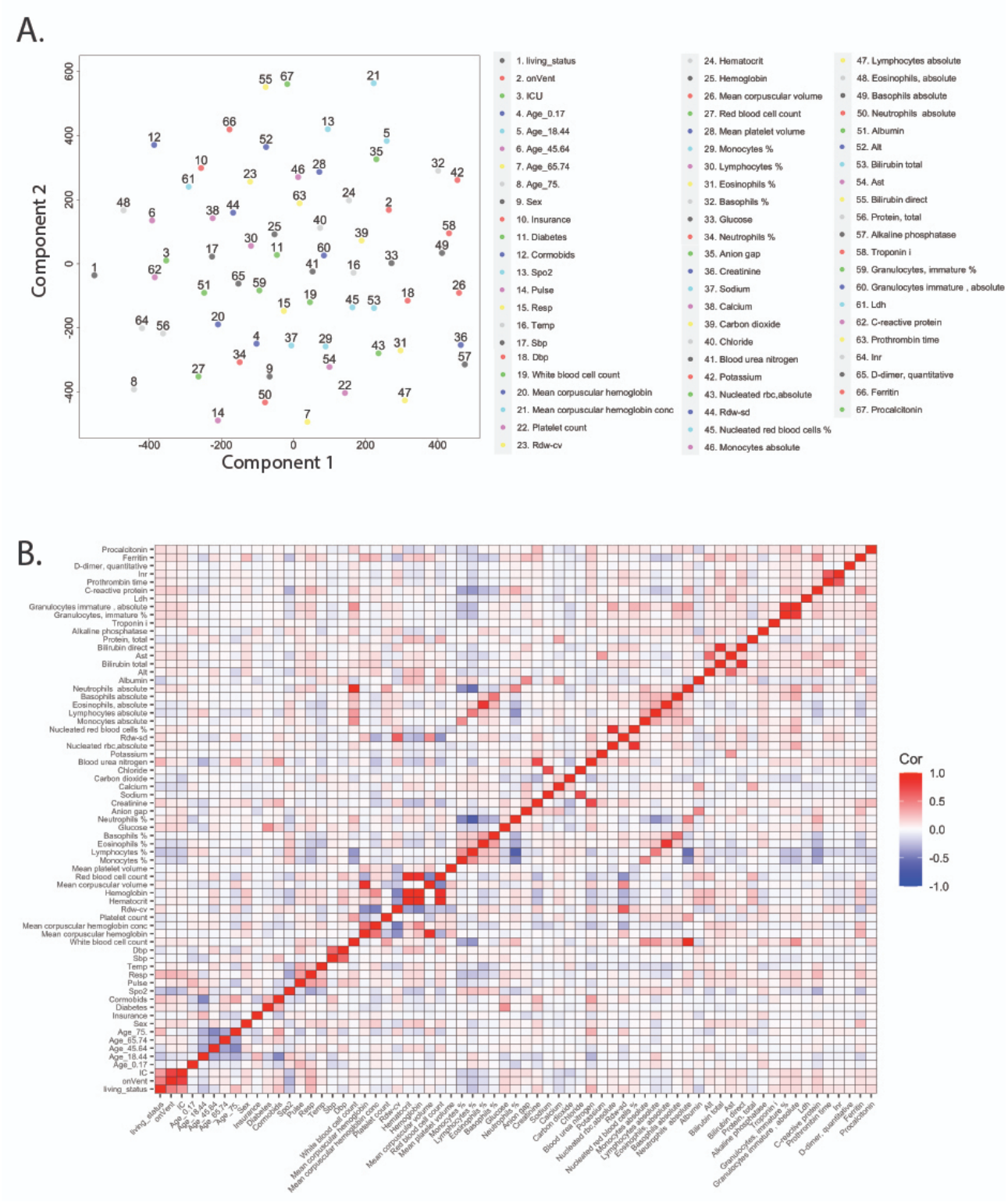
Figure 2. Predictive performance using clinical data from the first 24 hours. A. Receiver operating characteristic (ROC) curve and precision-recall curve (PRC) for logistic regression model. B. ROC curve and PRC for XGBoost (eXtreme Gradient Boosting) model. C. Coefficient weights for the logistic model are recorded on the left. Model performance gains for XGBoost are listed on the right. AUC: area under the curve; AUPRC: area under the precision-recall curve; Dbp: diastolic blood pressure; ICU: intensive care unit; Resp: respiration rate; Sbp: systolic blood pressure; SpO_2: peripheral oxygen saturation; Temp: temperature.



The 50 most frequently collected lab values and their relative importance were also studied. A t-distributed stochastic neighbor embedding plot (Figure 3A) suggests lack of clustering among lab features and overall low correlation (Figure 3B) in pairwise comparisons ($|\mu|=0.08$; $|\sigma|=0.10$). Local pockets of correlation ($|r|\geq 0.83$) were identified between hemoglobin, hematocrit, and red blood cell count; absolute neutrophils and white blood cell count; and bilirubin direct and bilirubin total. Each of these sets measures variables that are clinically interdependent and, thus, expected.

XSL•FO
RenderX

Figure 3. Overview of lab features collected in the first 24 hours. A. A t-distributed stochastic neighbor embedding plot based on previously collected clinical features and new lab values. B. Pairwise Pearson correlation heat map. Alt: alanine aminotransferase; Ast: aspartate aminotransferase; Dbp: diastolic blood pressure; ICU: intensive care unit; Inr: international normalized ratio; Ldh: lactate dehydrogenase; rbc: red blood cell; Resp: respiration rate; Rdw-cv: red cell distribution width–coefficient of variation; Rdw-sd: red cell distribution width–standard deviation; Resp: respiration rate; Sbp: systolic blood pressure; SpO_2: peripheral oxygen saturation; Temp: temperature.

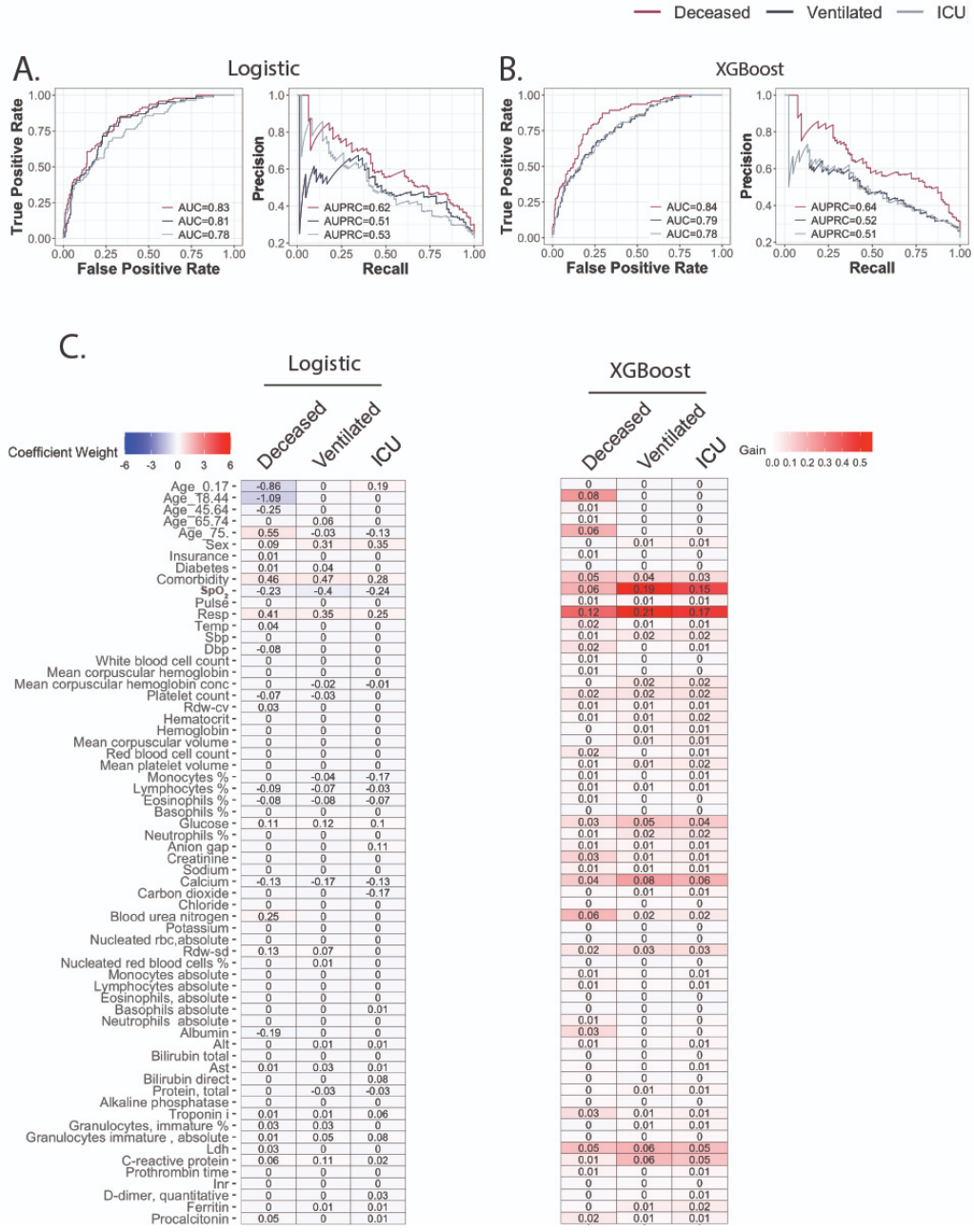


Incorporating lab features into the predictive models marginally improved performance. Logistic regression had AUC performances of 0.83, 0.81, and 0.78; specificities of 68%, 70%, and 69%; and sensitivities of 85%, 83%, and 74%, respectively (Figure 4A). The XGBoost model performed better, with AUC increasing to 0.84, 0.79, and 0.78; specificities of 71%, 72%, and 65%; and sensitivities of 83%, 73%, and 78%, respectively (Figure 4B). For logistic regression, blood urea nitrogen (BUN) and albumin were among the lab features (Figure 4C) that were predictive of mortality. The XGBoost model found the most performance gain from BUN and LDH. Feature importance for predicting ventilation or ICU admission differed between

XSL•FO
RenderX

models. For ventilation, logistic regression selected calcium, glucose, and CRP with large absolute coefficient values, while XGBoost identified calcium, glucose, CRP, and LDH as important features. For those admitted to ICU, XGBoost benefited from the same lab features, while monocyte percentage and carbon dioxide were additionally selected for by logistic regression. Of note, for XGBoost, no lab feature showed a higher importance measure than did respiration rate and SpO_2 for all three outcomes.

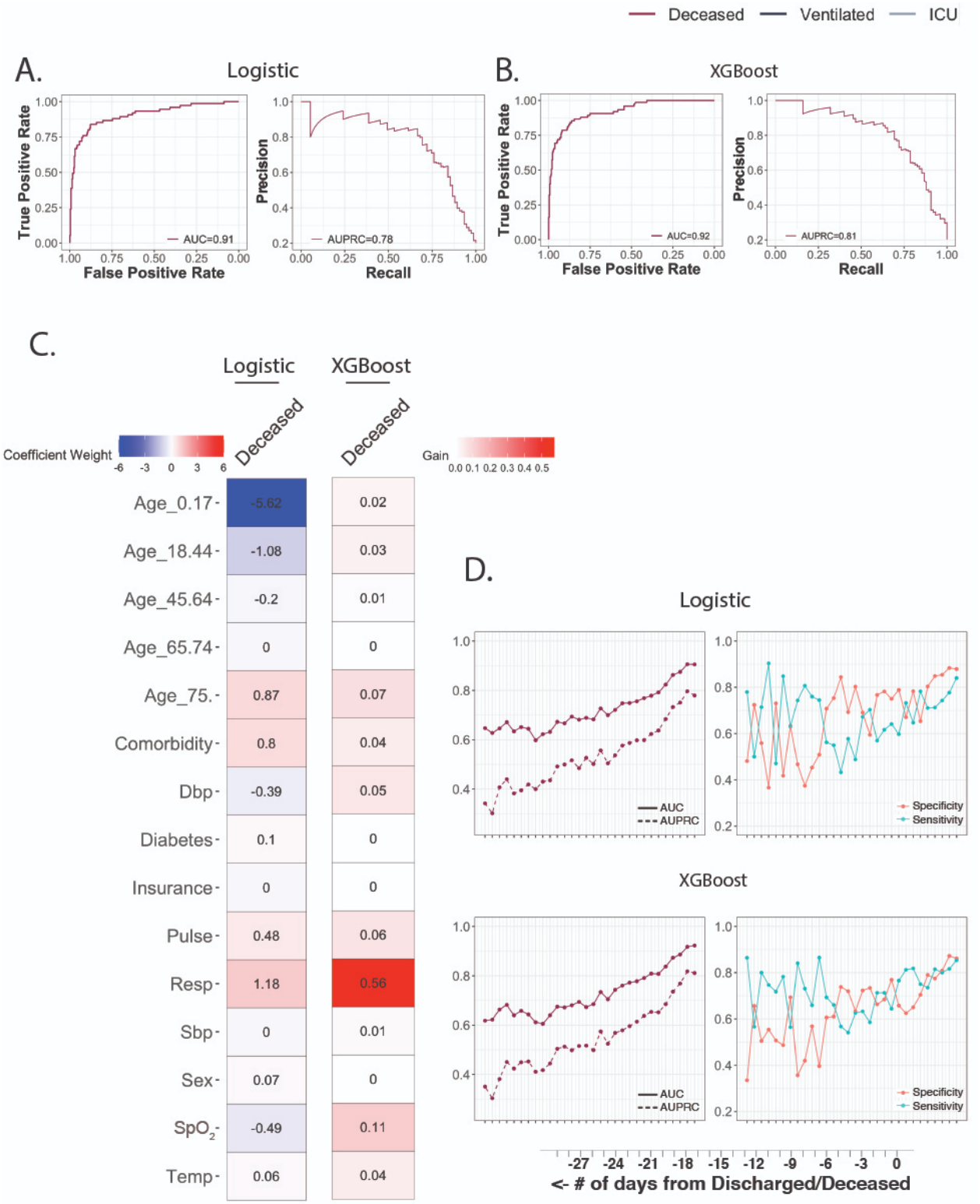
Figure 4. Predictive performance after incorporating lab features. A. Receiver operating characteristic (ROC) curve and precision-recall curve (PRC) for logistic regression model. B. ROC curve and PRC for the XGBoost (eXtreme Gradient Boosting) model. C. Coefficient weights for the logistic model are recorded on the left. Model performance gains for XGBoost are listed on the right. Alt: alanine aminotransferase; Ast: aspartate aminotransferase; AUC: area under the curve; AUPRC: area under the precision-recall curve; Dbp: diastolic blood pressure; ICU: intensive care unit; Inr: international normalized ratio; Ldh: lactate dehydrogenase; rbc: red blood cell; Resp: respiration rate; Rdw-cv: red cell distribution width–coefficient of variation; Rdw-sd: Red cell distribution width–standard deviation; Sbp: systolic blood pressure; SpO_2: peripheral oxygen saturation; Temp: temperature.



Finally, models trained on data collected in the last 24 hours excelled at predicting which patients would become deceased. The logistic regression model (Figure 5A) had an AUC performance of 0.91, specificity of 88%, and sensitivity of 84%. The XGBoost model (Figure 5B) had an AUC performance of 0.92, specificity of 86%, and sensitivity of 85%. The importance of respiration rate increased for XGBoost (Figure 5C), accounting for more than 50% of the gain. Values of SpO_2 and being aged 75 years and over were the next most important features.

XSL•FO
RenderX

Figure 5. Predictive performance of deceased using clinical data from the final 24 hours. A. Receiver operating characteristic (ROC) curve and precision-recall curve (PRC) for the logistic regression model. B. ROC curve and PRC for the XGBoost (eXtreme Gradient Boosting) model. C. Coefficient weights for the logistic model are recorded on the left. Model performance gains for XGBoost are listed on the right. D. Performance of models to predict deceased outcome was assessed using clinical data from the preceding 30 days. Plots track the area under the curve (AUC), area under the precision-recall curve (AUPRC), specificity, and sensitivity when using the threshold that maximized the sum of the sensitivity and specificity (Youden's J statistic). Dbp: diastolic blood pressure; ICU: intensive care unit; Resp: respiration rate; Sbp: systolic blood pressure; SpO_2: peripheral oxygen saturation; Temp: temperature.



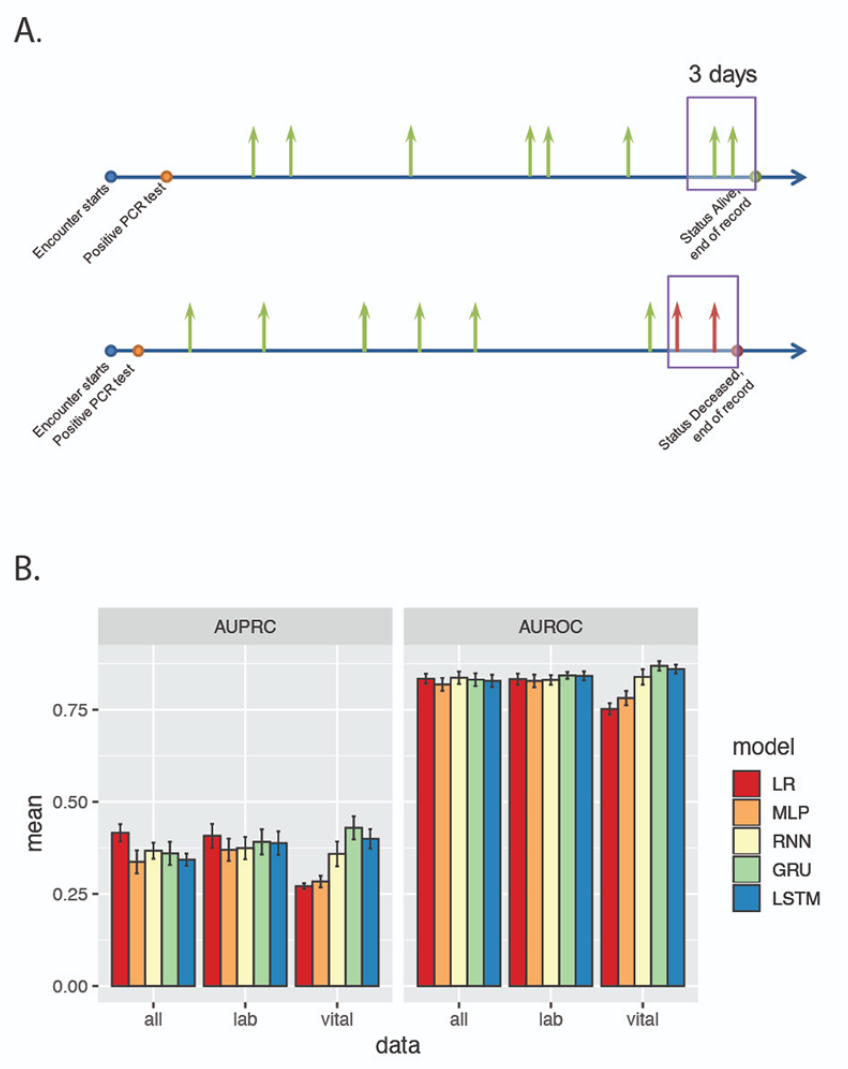
Using the same coefficients and tree weights and structures, both models were assessed based on clinical data from the preceding 30 days (Figure 5D). With cutoffs of 0.80 for AUC, and 70% for specificity and sensitivity, logistic regression was able to predict a deceased outcome 4 days in advance (AUC=0.82; specificity=85%; and sensitivity=71%) and 5 days in advance (AUC=0.81; specificity=70%; and sensitivity=75%) for XGBoost. Models were not trained on those ventilated or

XSL•FO
RenderX

ICU admitted, as these events are unlikely to occur in the final 24 hours preceding discharge and death. Lab values were not incorporated because few blood tests were ordered in the final 24 hours.

To explore whether patient status can be dynamically predicted based on past medical data, we also built time series models using simple RNN, GRU, and LSTM architectures and compared the performance metrics to single time point models of logistic regression and MLP. A major goal is to explore whether more complex modeling approaches are able to make accurate and transferrable predictions. The vital status of each patient was converted to a time series that was marked as positive if the time point was within 3 days of the patient becoming deceased (Figure 6A). The time series models (ie, RNN, GRU, and LSTM) were trained to take medical data from eight time points as input and infer the vital status at each point, giving rise to a real-time risk prediction based on historical records. The single time point models (ie, logistic regression and MLP), on the other hand, only used medical data at the current point to make the prediction. Model comparison was carried out with three different feature sets: vital signs only (ie, body temperature, pulse, respiration rate, systolic blood pressure, diastolic blood pressure, and SpO_2), vital signs and 46 lab results with nonzero coefficients in the single time point LASSO regression model, and vital signs and lab results plus *static* demographic information (ie, sex, age group, diabetic history, and comorbidities) (Multimedia Appendix 5). As the time series data were recorded at uneven and irregular intervals, the progression time (in days) was included in all models as an additional feature. For models only including vital sign features, time series models showed better performance (Figure 6B) compared to single time point models, but performance was comparable among all models when lab results and demographic information was added to the feature set.

Figure 6. Time series model performance. A. Time series labels for mortality risk. Medical measurements are obtained at time points across the admission period with uneven intervals. Green and red arrows represent time points with negative and positive (0/1) labels. B. Mean values of the area under the precision-recall curve (AUPRC) and area under the receiver operating characteristics curve (AUROC) for five model architectures across three feature settings. GRU: gated recurrent unit; LR: logistic regression; LSTM: long short-term memory; MLP: multilayer perceptron; PCR: polymerase chain reaction; RNN: recurrent neural network.



XSL•FO
RenderX

Discussion

Retrospective analysis of patients who tested positive for COVID-19 identified recognizable clinical markers, such as respiration rate and SpO_2, but also provided insights distinguishing morbidity (ie, ICU-admitted or ventilated outcomes) from mortality (ie, deceased outcome). Our study confirmed canonical risk factors that were previously established (eg, age, respiration rate, and SpO_2) as predictive of mortality and morbidity, but also uncovered the surprising finding that temperature was not predictive for mortality. In addition, lab markers of physiological stress, including LDH, BUN, and CRP, were found to be important for model prediction, but other canonical indicators, such as procalcitonin, were not.

Our results aligned with previous work [10] analyzing patient data from NYU Langone Health to predict absence of adverse events within a 96-hour window as opposed to negative outcomes. Several features of importance overlapped both studies, notably respiration rate, SpO_2, LDH, BUN, and CRP. However, other selected features, such as temperature, platelet count, pulse, and eosinophil percentage, were not found to be important in our model.

Although the goal of stratifying patients by disease severity aligned, the different approaches likely explain the differences in variable explanation. Our study differs in that our models were trained only on clinical data from the first 24 hours after admission, as compared to continuously updating predictions when new lab values were reported. Thus, features that are important for outcome prediction at the time of admission will differ from those that do a better job of modeling variations in disease severity over time. In addition, we stratified our negative outcomes into mortality and morbidity, and separated morbidity further to compare those requiring ICU admission versus ventilation. Eosinophil percentage was statistically different between all three clinical outcomes, while temperature and pulse were only different for morbidity and platelet counts were only different for mortality (Multimedia Appendix 2). It is hypothesized that patients exhibiting symptoms of fever and increased pulse rate, likely a consequence of decreased SpO_2 (r=–0.21 and –0.12, respectively), will likely be prioritized for ICU care and/or ventilation. Although SpO_2 and respiration rate were consistently selected as predictive features across outcomes and modeling methods, age groups were informative predictors of mortality risk only. As expected, the mortality model performed better than the morbidity models. These results suggest that disease severity and mortality risks may require unique modeling with different predictor subsets and weighting factors. It is also consistent with the observation that senior patients were the most vulnerable population, while the mortality rate among the youth was relatively low [11].

In addition, although current evidence suggests that adults with type 2 diabetes mellitus are at increased risk for COVID-19 complications, our XGBoost model did not find a past diagnosis important for predicting morbidity or mortality. Only after incorporating lab features did we identify a positive correlation between exact glucose values and poorer outcomes. Together, this observation suggests that the elevated blood sugar levels observed may be the result of physiological stress triggered by the disease. Indeed, prior work has shown that even when controlled for pre-existing diabetes, hyperglycemia was commonly observed in acutely ill hospitalized patients and linked to poorer outcomes [12,13].

Other lab features also identified routine chemistry data points that shed light on disease pathology. Values of LDH were elevated for all three clinical outcomes, a finding consistent with widespread tissue damage that has been shown in numerous studies to be a predictor of morbidity and mortality in a wide variety of diseases beyond COVID-19 [14-18]. Mortality was also predicted for by BUN. To investigate further the possibility of any relationship to acute kidney injury, we retrained our models with a BUN to creatinine ratio as an additional feature. While correlated with mortality (r=0.17), the feature was not selected for by LASSO, and was only of importance when BUN was removed from the training data set. Indeed, recent literature has revealed that BUN is emerging as an independent predictor of mortality in a variety of diseases, including heart failure [19], aortic dissection [20], and acute pancreatitis [21]. It has also been proposed that BUN is an important indicator for metabolic diseases and general nutritional status of patients, explaining its relative importance in the prediction for mortality. The relationship here is unclear and warrants further investigation.

Interestingly, calcium level upon admission was a more important predictor of morbidity in our models than procalcitonin was. As a peptide precursor of calcitonin, a hormone involved in calcium homeostasis, procalcitonin is also an acute phase reactant that has been used historically, albeit controversially, to help diagnose bacterial pneumonia [22-24]. Although many studies [25-27] have described a positive relationship between procalcitonin levels and mortality and morbidity in patients with COVID-19, few have commented on the importance of calcium as a prognostic value, as we have found in our study. Calcium was negatively correlated with all three measured clinical outcomes, which is consistent with other studies linking hypocalcemia with increased morbidity and mortality in patients with COVID-19 [28-30]. Theoretically, hypocalcemia could be a result of increased procalcitonin, since procalcitonin is the precursor of calcitonin whose function is to reduce serum calcium. Interestingly, it has been reported that in a systemic inflammatory response, serum calcitonin does not increase concordantly in response to increased procalcitonin. This situation could indicate that calcium is a predictive factor through an entirely different mechanism than the more well-established procalcitonin. One theory is that alteration of calcium homeostasis is perhaps used as a strategy by the SARS-CoV-2 virus for survival and replication, since calcium is essential for virus structure formation, entry, gene expression, virion maturation, and release. Another possibility is that patients who present with hypocalcemia have pre-existing parathyroid hormone (PTH) and vitamin D imbalances that are exacerbated by SARS-CoV-2 infection. Our study could not evaluate the importance of PTH or vitamin D due to infrequent lab orders (0.21% and 0.08% completeness, respectively).

While the inclusion of lab features resulted in only modest improvement for ventilation and ICU admission prediction, lab values did result in larger increases in performance metrics for

XSL•FO
RenderX

mortality prediction. However, time series modeling failed to improve prediction performance with more clinical features. This observation is likely due to the fact that laboratory results were sampled much less frequently than vital sign readings. Moreover, treating *static* demographic information as repeating time series measurements may be suboptimal for recurrent models. As discussed above, laboratory measurements may help in modeling mortality risk of patients, and future work will focus on efficiently incorporating these static features for dynamic predictions [31,32].

A key limitation of our data set revolves around balancing inclusion criteria to maximize the number of encounters available for model training, while also limiting the amount of missing data. For example, patients who test positive for COVID-19 and present with less severe symptoms in outpatient or telehealth settings may not have a complete set of vital signs or any lab values available. Similarly, past medical histories are dependent on an accurate recollection on the patient's part, either through a past hospital encounter or at the time of admission. In addition, it is possible that the comorbidities designation in our data set may have false negatives. Because patient histories are often self-reported, it is possible that admitted patients with no prior encounters with the hospital, or either the physical or cognitive inability to verbalize such history at the time of triage, would not have such indication available in the electronic health record. However, this reflects real-world medical situations, in which diagnoses must be made based on unverifiable patient data or delayed lab results. Finally, as data were retrospectively gathered from Epic during the early stages of the pandemic, when diagnostic and treatment protocols were still being developed, a concerted effort to gather novel biomarker tests that have later been shown to be linked with disease severity is not expected. As time draws on and new variants emerge, we also expect that repeated studies will be needed to survey changes to risk factors.

Acknowledgments

We wish to thank MCIT and the Office of Science and Research at NYU Langone Health for maintaining and deidentifying the clinical database. JMW is supported by the NYU Medical Scientist Training Program (T32GM136573); he is also supported, in part, by an NYU Clinical and Translational Science Institute grant (TL1 TR001447) from the National Center for Advancing Translational Sciences, National Institutes of Health. The content in this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. A portion of this work was also funded by Renaissance Health Service Corporation and Delta Dental of Michigan.

Authors' Contributions

JMW, WL, and DF contributed to the conception and design of the study; contributed to data acquisition, analysis, and interpretation; and drafted and critically revised the manuscript. XC contributed to the conception and design of the study, contributed to data analysis and interpretation, and drafted and critically revised the manuscript. MPM and JTM contributed to the conception and design of the study, contributed to data interpretation, and critically revised the manuscript.

Conflicts of Interest

MPM has served as a paid consultant for SensoDx, LLC, and has a provisional patent pending. JTM has a provisional patent pending; in addition, he has an ownership position and an equity interest in both SensoDx, LLC, and OraLiva, Inc, and serves on their advisory boards.

Multimedia Appendix 1
Flow diagram to illustrate the number of patients at each filtering step and descriptions of the criteria that needed to be fulfilled.
[PDF File (Adobe PDF File), 59 KB - jmir_v23i7e29514_app1.pdf]

Multimedia Appendix 2
Demographics, vital signs, lab results, and clinical outcomes of the retrospective cohort.
[DOCX File , 56 KB - jmir_v23i7e29514_app2.docx]

Multimedia Appendix 3
Effect of early ventilation treatment on vital signs. A. Distribution of averaged respiration rates for all values preceding 24 hours (top) and for all values that precede initiation of ventilation treatment (bottom). B. Distribution of averaged SpO₂ (peripheral oxygen saturation) for all values preceding 24 hours (top) and for all values that precede initiation of ventilation treatment (bottom).
[PDF File (Adobe PDF File), 1021 KB - jmir_v23i7e29514_app3.pdf]

Multimedia Appendix 4
XGBoost (eXtreme Gradient Boosting) parameter settings for each trained model.
[DOCX File , 17 KB - jmir_v23i7e29514_app4.docx]

XSL•FO
RenderX

Multimedia Appendix 5
Description of features utilized for time series modeling.
[DOCX File , 22 KB - jmir_v23i7e29514_app5.docx]

References

1. COVID Data Tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker [accessed 2020-12-09]
2. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
3. Yao H, Zhang N, Zhang R, Duan M, Xie T, Pan J, et al. Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests. Front Cell Dev Biol 2020;8:683 [FREE Full text] [doi: 10.3389/fcell.2020.00683] [Medline: 32850809]
4. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
5. Cheng F, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med 2020 Jun 01;9(6):1668 [FREE Full text] [doi: 10.3390/jcm9061668] [Medline: 32492874]
6. McRae M, Dapkins I, Sharif I, Anderman J, Fenyo D, Sinokrot O, et al. Managing COVID-19 with a clinical decision support tool in a community health network: Algorithm development and validation. J Med Internet Res 2020 Aug 24;22(8):e22033 [FREE Full text] [doi: 10.2196/22033] [Medline: 32750010]
7. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 1996;58(1):267-288 [FREE Full text] [doi: 10.1111/j.2517-6161.1996.tb02080.x]
8. Chen T, Guestrin C. XGBoost: A scalable rree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: Association for Computing Machinery; 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
9. Keras Team. Keras. GitHub. 2020. URL: https://github.com/keras-team/keras [accessed 2020-12-09]
10. Razavian N, Major VJ, Sudarshan M, Burk-Rafel J, Stella P, Randhawa H, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med 2020;3:130 [FREE Full text] [doi: 10.1038/s41746-020-00343-x] [Medline: 33083565]
11. Yang W, Kandula S, Huynh M, Greene S, Van Wye G, Li W, et al. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: A model-based analysis. Lancet Infect Dis 2021 Feb;21(2):203-212 [FREE Full text] [doi: 10.1016/S1473-3099(20)30769-6] [Medline: 33091374]
12. Corathers SD, Falciglia M. The role of hyperglycemia in acute illness: Supporting evidence and its limitations. Nutrition 2011 Mar;27(3):276-281. [doi: 10.1016/j.nut.2010.07.013] [Medline: 20869205]
13. McCowen KC, Malhotra A, Bistrian BR. Stress-induced hyperglycemia. Crit Care Clin 2001 Jan;17(1):107-124. [doi: 10.1016/s0749-0704(05)70154-8] [Medline: 11219223]
14. Kishaba T, Tamaki H, Shimaoka Y, Fukuyama H, Yamashiro S. Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. Lung 2014 Feb;192(1):141-149. [doi: 10.1007/s00408-013-9530-0] [Medline: 24221341]
15. Lam CWK, Chan MHM, Wong CK. Severe acute respiratory syndrome: Clinical and laboratory manifestations. Clin Biochem Rev 2004 May;25(2):121-132 [FREE Full text] [Medline: 18458712]
16. Chen X, Huang M, Xiao Z, Yang S, Chen X. Lactate dehydrogenase elevations is associated with severity of COVID-19: A meta-analysis. Crit Care 2020 Jul 24;24(1):459 [FREE Full text] [doi: 10.1186/s13054-020-03161-5] [Medline: 32709251]
17. LitCovid. URL: https://www.ncbi.nlm.nih.gov/research/coronavirus/ [accessed 2020-12-09]
18. Goldberg D, Brown D. Biochemical tests in the diagnosis, classification, and management of patients with malignant lymphoma and leukemia. Clin Chim Acta 1987 Oct;169(1):1-76. [doi: 10.1016/0009-8981(87)90394-9]
19. Filippatos G, Rossi J, Lloyd-Jones DM, Stough WG, Ouyang J, Shin DD, et al. Prognostic value of blood urea nitrogen in patients hospitalized with worsening heart failure: Insights from the Acute and Chronic Therapeutic Impact of a Vasopressin Antagonist in Chronic Heart Failure (ACTIV in CHF) study. J Card Fail 2007 Jun;13(5):360-364. [doi: 10.1016/j.cardfail.2007.02.005] [Medline: 17602982]
20. Liu J, Sun L, Wang J, Ji G. Blood urea nitrogen in the prediction of in-hospital mortality of patients with acute aortic dissection. Cardiol J 2018;25(3):371-376 [FREE Full text] [doi: 10.5603/CJ.a2017.0075] [Medline: 28653311]
21. Zhou H, Mei X, He X, Lan T, Guo S. Severity stratification and prognostic prediction of patients with acute pancreatitis at early phase: A retrospective study. Medicine (Baltimore) 2019 Apr;98(16):e15275 [FREE Full text] [doi: 10.1097/MD.0000000000015275] [Medline: 31008971]
22. Wussler D, Kozhuharov N, Tavares Oliveira M, Bossa A, Sabti Z, Nowak A, et al. Clinical utility of procalcitonin in the diagnosis of pneumonia. Clin Chem 2019 Dec;65(12):1532-1542. [doi: 10.1373/clinchem.2019.306787] [Medline: 31615771]

XSL•FO
RenderX

23. Peck Palmer OMG. Procalcitonin's adjunct role in the diagnosis and management of pneumonia. Clin Chem 2019 Dec;65(12):1474-1476. [doi: 10.1373/clinchem.2019.312520] [Medline: 31672860]
24. Self WH, Balk RA, Grijalva CG, Williams DJ, Zhu Y, Anderson EJ, et al. Procalcitonin as a marker of etiology in adults hospitalized with community-acquired pneumonia. Clin Infect Dis 2017 Jul 15;65(2):183-190 [FREE Full text] [doi: 10.1093/cid/cix317] [Medline: 28407054]
25. Liu Z, Li J, Wang S, Chen D, Zeng W, Chen S, et al. Association of procalcitonin levels with the progression and prognosis of hospitalized patients with COVID-19. Int J Med Sci 2020;17(16):2468-2476 [FREE Full text] [doi: 10.7150/ijms.48396] [Medline: 33029089]
26. Hu R, Han C, Pei S, Yin M, Chen X. Procalcitonin levels in COVID-19 patients. Int J Antimicrob Agents 2020 Aug;56(2):106051 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.106051] [Medline: 32534186]
27. Liu F, Li L, Xu M, Wu J, Luo D, Zhu Y, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. J Clin Virol 2020 Jun;127:104370 [FREE Full text] [doi: 10.1016/j.jcv.2020.104370] [Medline: 32344321]
28. Liu J, Han P, Wu J, Gong J, Tian D. Prevalence and predictive value of hypocalcemia in severe COVID-19 patients. J Infect Public Health 2020 Sep;13(9):1224-1228 [FREE Full text] [doi: 10.1016/j.jiph.2020.05.029] [Medline: 32622796]
29. Sun J, Zhang W, Zou L, Liu Y, Li J, Kan X, et al. Serum calcium as a biomarker of clinical severity and prognosis in patients with coronavirus disease 2019. Aging (Albany NY) 2020 Jun 25;12(12):11287-11295 [FREE Full text] [doi: 10.18632/aging.103526] [Medline: 32589164]
30. Cappellini F, Brivio R, Casati M, Cavallero A, Contro E, Brambilla P. Low levels of total and ionized calcium in blood of COVID-19 patients. Clin Chem Lab Med 2020 Aug 27;58(9):e171-e173. [doi: 10.1515/cclm-2020-0611] [Medline: 32459190]
31. Miebs G, Mochol-Grzelak M, Karaszewski A, Bachorz RA. Efficient strategies of static features incorporation into the recurrent neural network. Neural Process Lett 2020 Jan 31;51(3):2301-2316. [doi: 10.1007/s11063-020-10195-x]
32. Esteban C, Staeck O, Baier S, Yang Y, Tresp V. Predicting clinical events by combining static and dynamic information using recurrent neural networks. In: Proceedings of the 2016 IEEE International Conference on Healthcare Informatics. 2016 Presented at: 2016 IEEE International Conference on Healthcare Informatics; October 4-7, 2016; Chicago, IL p. 93-101. [doi: 10.1109/ICHI.2016.16]

Abbreviations

AUC: area under the curve
BUN: blood urea nitrogen
CRP: C-reactive protein
GRU: gated recurrent unit
HIPAA: Health Insurance Portability and Accountability Act
ICU: intensive care unit
IRB: Institutional Review Board
LASSO: least absolute shrinkage and selection operator
LDH: lactate dehydrogenase
LSTM: long short-term memory
MCIT: Medical Center Information Technology
MLP: multilayer perceptron
NYU: New York University
PCR: polymerase chain reaction
PTH: parathyroid hormone
RNN: recurrent neural network
SpO_2: peripheral oxygen saturation
XGBoost: eXtreme Gradient Boosting

XSL•FO
RenderX

Edited by C Basch; submitted 09.04.21; peer-reviewed by E Mahmoudi, X Cheng; comments to author 02.05.21; revised version received 17.05.21; accepted 26.05.21; published 09.07.21.

Please cite as:
Wang JM, Liu W, Chen X, McRae MP, McDevitt JT, Fenyö D
Predictive Modeling of Morbidity and Mortality in Patients Hospitalized With COVID-19 and its Clinical Implications: Algorithm Development and Interpretation
J Med Internet Res 2021;23(7):e29514
URL: https://www.jmir.org/2021/7/e29514
doi:10.2196/29514
PMID:34081611

©Joshua M Wang, Wenke Liu, Xiaoshan Chen, Michael P McRae, John T McDevitt, David Fenyö. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Racial and Ethnic Differences in COVID-19 Outcomes, Stressors, Fear, and Prevention Behaviors Among US Women: Web-Based Cross-sectional Study

Jamila K Stockman[1*], MPH, PhD; Brittany A Wood[1*], MPH; Katherine M Anderson[1,2*], MPH

[1]Division of Infectious Diseases and Global Public Health, Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, CA, United States

[2]Department of Behavioral, Social, and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Jamila K Stockman, MPH, PhD
Division of Infectious Diseases and Global Public Health
Department of Medicine, School of Medicine
University of California, San Diego
9500 Gilman Drive, MC0507
La Jolla, CA, 92093-0507
United States
Phone: 1 8588224652
Fax: 1 8585347566
Email: jstockman@health.ucsd.edu

Abstract

Background: In the United States, racial and ethnic minorities are disproportionately affected by COVID-19, with persistent social and structural factors contributing to these disparities. At the intersection of race/ethnicity and gender, women of color may be disadvantaged in terms of COVID-19 outcomes due to their role as essential workers, their higher prevalence of pre-existing conditions, their increased stress and anxiety from the loss of wages and caregiving, and domestic violence.

Objective: The purpose of this study is to examine racial and ethnic differences in the prevalence of COVID-19 outcomes, stressors, fear, and prevention behaviors among adult women residing in the United States.

Methods: Between May and June 2020, women were recruited into the Capturing Women's Experiences in Outbreak and Pandemic Environments (COPE) Study, a web-based cross-sectional study, using advertisements on Facebook; 491 eligible women completed a self-administered internet-based cross-sectional survey. Descriptive statistics were used to examine racial and ethnic differences (White; Asian; Native Hawaiian or other Pacific Islander; Black; Hispanic, Latina, or Spanish Origin; American Indian or Alaskan Native; multiracial or some other race, ethnicity, or origin) on COVID-19 outcomes, stressors, fear, and prevention behaviors.

Results: Among our sample of women, 16% (73/470) reported COVID-19 symptoms, 22% (18/82) were concerned about possible exposure from the people they knew who tested positive for COVID-19, and 51.4% (227/442) knew where to get tested; yet, only 5.8% (27/469) had been tested. Racial/ethnic differences were observed, with racial/ethnic minority women being less likely to know where to get tested. Significant differences in race/ethnicity were observed for select stressors (food insecurity, not enough money, homeschooling children, unable to have a doctor or telemedicine appointment) and prevention behaviors (handwashing with soap, self-isolation if sick, public glove use, not leaving home for any activities). Although no racial/ethnic differences emerged from the Fear of COVID-19 Scale, significant racial/ethnic differences were observed for some of the individual scale items (eg, being afraid of getting COVID-19, sleep loss, and heart racing due to worrying about COVID-19).

Conclusions: The low prevalence of COVID-19 testing and knowledge of where to get tested indicate a critical need to expand testing for women in the United States, particularly among racial/ethnic minority women. Although the overall prevalence of engagement in prevention behaviors was high, targeted education and promotion of prevention activities are warranted in communities of color, particularly with consideration for stressors and adverse mental health.

(J Med Internet Res 2021;23(7):e26296) doi:10.2196/26296

XSL•FO
RenderX

KEYWORDS

COVID-19; race; ethnicity; demographics; stress; fear; prevention behaviors; prevalence; cross-sectional; women; United States

Introduction

The COVID-19 pandemic has shed light on racial and ethnic disparities in the United States [1-3]. Among 44% of US cases with known race or ethnicity, 33% were Latinx, 22% were Black, and 1.3% were American Indian or Alaskan Native (AIAN), despite accounting for 18%, 13%, and 0.7% of the US population, respectively, suggesting these groups are disproportionately affected by the COVID-19 pandemic [1]. Black individuals are almost three times more likely to be hospitalized for COVID-19 than White individuals [4]. Nationally, COVID-19 mortality is 80% higher for Black individuals and over 50% higher for Latinx individuals, relative to White individuals [5].

Contributing factors to racial and ethnic disparities in COVID-19 outcomes include social and structural vulnerabilities and pre-existing health conditions, affecting socially marginalized populations [6,7]. Residing in high-poverty areas, being underinsured or uninsured, and having limited access to health care likely contribute to limited COVID-19 testing, resulting in late presentation with more advanced disease [8-10]. Historically, Black individuals and AIAN individuals have endured systemic racism and discrimination, a source of structural inequities [10], that has created deep mistrust of medical institutions and unequal access to care and treatment [11]. Ethnic minority groups are overrepresented in essential service industries, limiting the ability to shelter at home and increasing the likelihood of low wages, unpaid sick leave, or limited workplace protections [10,12]. The physical environment also may increase COVID-19 risk through residential segregation, residing in densely populated areas, and multigenerational households; the resulting physical crowding contributes to increased viral transmission [7,13]. Ethnic minority groups are disproportionately affected by chronic medical conditions, which may worsen COVID-19 outcomes [2]. Moreover, toxic stress resulting from racial and social inequities has been magnified during the pandemic, with implications for poor biological function and adverse physical and mental health outcomes [11,14]. This culminates in Black, Latinx, and AIAN groups shouldering greater disease burden and the unfolding spread of COVID-19 in areas with larger populations of ethnic minorities [8].

The continuation of the COVID-19 pandemic has led the US government to recommend prevention behaviors, including staying at home except for essential activities, physical distancing from nonhousehold members, wearing a face mask in public, avoiding nonhousehold groups, and frequent handwashing [15]. Compliance with these recommendations has been variable [16-19]. A May 2020 Centers for Disease Control and Prevention (CDC) survey found that 91.5% of participants had been to a public area in the previous week, 77.3% stayed home except for essential activities, and 75.1% and 10.8% were “always” and “often” avoiding groups of 10 or more people, respectively [17]. Further, physical distancing was maintained “always” by 58.2% and “often” by 21.3% of respondents, and 60.3% “always” and 13.8% “often” wore a face mask in public [17]. A survey conducted in early April 2020 found that 87% of participants were physically distancing in public, while only 50% used face masks in public; older age, female gender, and financial security were associated with adherence, with no differences by race or ethnicity, possibly due to overrepresentation of White participants [16]. Other assessments of racial and ethnic differences in prevention vary. In one study, Black adults knew less about symptoms and routes of transmission of COVID-19, left their homes more frequently, and handwashed less often than White adults [19]; in another, Black adults were less likely to maintain physical distance or avoid groups compared to their White counterparts but more likely to stay home except for essential activities, more likely to “always” or “often” wear a face mask in public, and equally likely to have been in a public area in the preceding week [17].

Given the lack of published findings on COVID-19 outcomes by race/ethnicity among women in the United States, our paper describes the prevalence and racial/ethnic differences among COVID-19 outcomes, stressors, fear, and prevention behaviors among an ethnically diverse sample of US adult women.

Methods

Study Population

Between May and June 2020, we conducted the Capturing Women’s Experiences in Outbreak and Pandemic Environments (COPE) Study, a web-based cross-sectional study of COVID-19 outcomes, testing, and prevention behaviors among adult women in the United States. Our study included participants who were 18 years or older, cisgender or transgender female, living in the United States, and able to understand English. We excluded participants who did not reside within the United States, were younger than 18 years, were male, or did not understand English.

Recruitment and Enrollment

Recruitment occurred through Facebook advertisements, designed to recruit adult women currently residing in the United States. We aimed to collect at least 400 survey responses. Interested women accessed a digital consent form within the REDCap (Research Electronic Data Capture) survey platform. The consent form included the purpose of the study and participant risk, burden, rights, and compensation. Participants who provided documented consent were directed to a screening survey. Screening questions included age, gender, and state of residence. Of 682 users who clicked on an advertisement, 633 (92.8%) provided informed consent and answered screener questions. Among the 633 participants, 7 participants were not eligible for the study. Eligible participants (n=626) were directed to the survey, which took 15 to 20 minutes to complete. Of 626 eligible participants, 491 (78% response rate) completed the full web-based survey. Participants received a US $20 Amazon e-gift card via email. The COPE Study procedures were approved by the University of California, San Diego Human Research Protections Program Institutional Review Board.

XSL•FO
RenderX

Survey Measures

Sociodemographic Characteristics

Race/ethnicity were self-reported by using the following categories: White (eg, German, Irish, English, Italian, Polish, or French); Asian, Native Hawaiian, or other Pacific Islander (API); Black (eg, African American or Black Caribbean); Hispanic, Latina, or Spanish origin (eg, Mexican or Mexican American, Puerto Rican, Cuban, or Dominican); AIAN (eg, Navajo nation, Blackfeet tribe, or Mayan); and multiracial or some other race, ethnicity, or origin. "Other race, ethnicity, or origin" included 7 women who self-identified as Middle Eastern or North African. Other demographics included age, education, employment status, sexual orientation, relationship status, parenthood, household composition and type of occupants, type of residential community (eg, urban or rural), current living situation, and household income.

COVID-19–Related Outcomes

We modified COVID-19 measures developed by the World Health Organization (WHO) for rapid behavioral studies about COVID-19 [20]. We assessed whether participants and close others (eg, family, friend, partner, or coworker) had COVID-19 symptoms, were tested, knew where to get tested, were diagnosed, were hospitalized, or died from COVID-19.

COVID-19 Stressors and Fear

We assessed nine COVID-19 stressors as outlined by the WHO [20]. Examples included food insecurity, insufficient rent, and caregiver status. We employed the Fear of COVID-19 Scale, a 7-item self-reported measure of an individual's fear of COVID-19 [21]. Items include fear of dying of COVID-19 and loss of sleep due to worrying about COVID-19. Responses followed a 5-point Likert scale from "strongly disagree" to "strongly agree." Individual items were assessed, and the overall score was calculated by summing all 7 items (range 7-35; Cronbach alpha .90).

COVID-19 Prevention Behaviors

COVID-19 prevention behaviors examined were: washing hands, avoiding touching face, using disinfectants (including hand sanitizer), staying home except for essential activities, covering mouth when coughing, physical distancing from nonhousehold members, physical distancing if sick, isolating oneself if sick, public face mask use, public glove use, avoiding crowds, and not leaving home [20]. Responses to each were dichotomized.

Data Analysis

Excluding participants without reported race/ethnicity, the sample was 473 women. We computed descriptive statistics of all variables overall and by race/ethnicity. We report medians and IQRs for nonnormally distributed continuous variables and frequencies and proportions for categorical variables. For race/ethnicity comparisons, we used Kruskal-Wallis tests and chi-square tests. If subgroups had expected values less than 5, Fisher exact test or a Monte Carlo estimate for Fisher exact *P* value was used. Individual pairwise comparisons by race/ethnicity were conducted when the Kruskal-Wallis test or chi-square test had a $P<.05$. The Dwass-Steel-Critchlow-Fligner test was used when the Kruskal-Wallis test was significant ($P<.05$), and individual chi-square tests between each of the racial/ethnic groups were examined when the overall chi-square was significant ($P<.05$). To account for potential confounding, COVID-19–related outcomes, stressors, and prevention behaviors that were significantly associated in bivariate analyses were included in multivariable binary logistic regressions. Each regression was adjusted for age, education, income, and type of residential community (urban, suburban, rural), with race as the independent variable and the outcome, stressor, or prevention behavior as the dependent variables. All analyses were performed using SAS, version 9.4 (SAS Institute).

Results

Sociodemographics

Of 473 women, 51% (n=241) were White, 13.5% (n=64) API, 12.7% (n=60) Black or African American, 10.1% (n=48) Hispanic or Latinx, 5.7% (n=27) AIAN, and 7% (n=33) multiracial or "other" (Multimedia Appendix 1). The median age was 33 years, with White women significantly older than API, AIAN, and multiracial or other women. Most women had completed some college or graduate school (377/463, 79.7%). Approximately 40% (175/473) of women were unemployed. Approximately 70% (323/471) were in a relationship, and 45% (208/471) had children.

Approximately 46% (211/461) of women lived alone. Of 250 living with others, 55.2% (n=138) and 53.2% (n=133) lived with their partner and family members, respectively. The median number of children in the household was 2. Almost 44% (193/447) had a household income of US $50,000 or more, with White individuals more likely than all racial/ethnic groups, except Latinx. Almost 50% (212/466) of women lived in an urban environment and 17.6% (82/466) in a rural environment.

COVID-19 Outcomes

Approximately 16% (73/470) of women reported COVID-19 symptoms; the highest prevalence was among White (45/240, 18.8%) and multiracial or other (7/33, 21.2%) women, and the lowest prevalence was among API women (4/63, 6.4%; Multimedia Appendix 2). Although 51.4% (227/442) of women knew where to get tested for COVID-19, only 5.8% (27/469) had been tested. White women were more likely to know where to get tested than API, Latinx, and AIAN women ($P<.001$, $P=.01$, and $P=.01$, respectively), and Black women more likely than API and AIAN women ($P=.01$ and $P=.03$, respectively). A total of 2% (8/469) of women were diagnosed with COVID-19, of which 50% (4/8) were hospitalized; 22% (18/82) of women were concerned that they were exposed to COVID-19 by the people they know. No significant differences by race/ethnicity were found for either.

A total of 18% (82/469) of women knew someone diagnosed with COVID-19, with White and Black women more likely than API women ($P=.02$ for both); 67.1% (55/82) identified a diagnosed friend, and 42.7% (35/82) identified diagnosed family. Of the 82 women, 42.7% (n=35) knew someone hospitalized, with Black women more likely than White women ($P<.001$); family members (16/35, 45.7%) and friends (14/35,

XSL•FO
RenderX

40.0%) were the most frequently reported relationships. Black women were more likely than White women to report a family member hospitalized (*P*=.01). Of the 82 women, 22% (n=18) knew someone who had died of COVID-19; most were a friend (8/18, 44.4%) or family member (6/18, 33.3%).

COVID-19 Stressors and Fear

Racial/ethnic differences emerged for four COVID-19 stressors (Multimedia Appendix 3). A total of 17% (82/473) of women experienced food insecurity. White women were less likely than API, Black, Latinx, AIAN, and multiracial or other race women to not have enough food (*P*<.001, *P*<.001, *P*=.03, *P*<.001, and *P*=.002, respectively); API women were more likely than Latinx women (*P*=.046) to not have enough food. Approximately 20% (91/473) did not have enough money for rent; White women were less likely to not have enough money for rent than API, Black, Latinx, and AIAN women (*P*<.001, *P*<.001, *P*=.03, and *P*<.001, respectively). Of 20.9% (99/473) who reported homeschooling children, White women were more likely than API and AIAN women (*P*=.03 and *P*=.003, respectively), and Black and Latinx women were more likely than AIAN women (*P*=.004 and *P*=.01, respectively). Of 28.1% (133/473) unable to have a doctor or telemedicine appointment, White women were more likely than API and AIAN women (*P*=.003 and *P*=.03, respectively). For the remaining stressors, 36.6% (173/473) of women reported loss of income, 15.4% (73/473) were essential employees, 4.7% (22/473) were unable to obtain medication, and 9.9% (47/473) were caregivers for sick family without or with (15/473, 3.2%) COVID-19.

Although no racial/ethnic differences emerged for the Fear of COVID-19 Scale, significant differences were observed for individual items. Higher agreement with the statement "I am very afraid of coronavirus" was found among White women compared to API and multiracial or other women (*P*=.01 and *P*=.02, respectively). White women reported more agreement with becoming nervous or anxious when watching media about COVID-19 compared to API women (*P*=.001). Latinx women agreed with experiencing clammy hands when thinking about COVID-19 more than White and multiracial or other women (*P*=.02 for both); AIAN women agreed more than White, Black, and multiracial or other women (*P*=.002, *P*=.03, and *P*=.02, respectively). AIAN women agreed more with sleep loss due to worrying about getting COVID-19 compared to White, Black, and multiracial or other women (*P*=.01, *P*=.002, and *P*=.01, respectively); API women agreed more than Black women (*P*=.02). AIAN women also had more agreement with experiencing their heart racing when thinking about getting COVID-19 compared to Black women (*P*=.04).

COVID-19 Prevention Behaviors

Significant differences in race/ethnicity were observed for four COVID-19 prevention behaviors (Multimedia Appendix 4). Of 95.6% (452/473) of women reporting handwashing with soap, White and API women were more likely than AIAN women (*P*=.002 and *P*=.02, respectively). Of 60.0% (284/473) who self-isolated if sick, API women were more likely than White, Black, and multiracial or other women (*P*=.01, *P*=.004, and *P*=.002, respectively). Of 40.2% (190/473) who publicly used gloves, API, Black, and AIAN women were more likely than White women (*P*<.001, *P*=.047, and *P*<.001, respectively); API women were more likely to publicly use gloves than Latinx and multiracial or other women (*P*<.001 for both); AIAN women were more likely to publicly use gloves than Black, Latinx, and multiracial or other women (*P*=.02, *P*=.01, and *P*<.001, respectively). Of 43.6% (206/473) who did not leave home at all, White women were less likely than API, AIAN, and multiracial or other women (*P*<.001, *P*=.001, and *P*=.04, respectively); API women were more likely to not leave home at all than Black and Latinx women (*P*=.01 and *P*=.001, respectively); AIAN women were more likely to not leave home at all than Latinx women (*P*=.02). Approximately 80% (374/473) of women avoided face touching, 87.1% (412/473) used disinfectants, 88.6% (419/473) stayed home except for essential activities, 82.2% (389/473) covered mouths when coughing, 86.5% (409/473) physically distanced from nonhousehold members, 59.8% (283/473) physically distanced from others if sick, 79.7% (377/473) publicly used a face mask, and 81.8% (387/473) avoided crowds.

To account for potential confounding, COVID-19–related outcomes, stressors, and prevention behaviors that were significantly associated with race or ethnicity in bivariate analyses were identified for inclusion in multivariable binary logistic regressions. Each regression was adjusted for age, education, income, and type of residential community (urban, suburban, rural). Results of these regressions are available in Multimedia Appendix 5 and confirm that there are statistically significant differences in all COVID-19–related outcomes, stressors, and prevention behaviors by race and ethnicity in comparison to the reference group, after accounting for potential confounders.

Discussion

Principal Findings

To our knowledge, this is the first study to describe COVID-19 outcomes, stressors, fear, and prevention behaviors among a sample of US women and to examine racial/ethnic differences. Our data showed that 1 in 6 women had COVID-19 symptoms. Troublingly, although 51.4% (227/442) of women knew where to get tested for COVID-19 and 22.0% (18/82) were concerned about possible exposure to COVID-19, only 5.8% (27/469) had been tested.

White women were the most likely to know where to be tested for COVID-19, despite experiencing a lower burden of COVID-19 among family and friends compared to Black women. This may be due to persistent medical mistrust due to legacies of abuse, mistreatment [22,23], and discrimination as an antecedent to medical mistrust [24]. Social and economic inequalities shape and sustain medical mistrust, particularly among populations who experience staggering health disparities [23]. Further, this may be reflective of an inability to access testing due to structural inequities [10], compounded by employment as essential workers [10,12], limiting available sick leave for testing and support for accessing health care. Given this, it is imperative to engage underserved populations in education and testing through community mobilization, social network, and peer-led engagement approaches.

XSL•FO
RenderX

A total of 18% (82/469) of women knew someone close to them diagnosed with COVID-19—often family or friends—with White and Black women being more likely than API women. Black women were more likely than White women to know someone hospitalized for COVID-19 and report them as family. This demonstrates a high burden of COVID-19 morbidity and mortality among Black families and their social circles; further, this implies a potential burden of adverse mental health associated with loss and trauma.

Racial/ethnic differences were observed for select stressors. White women were less likely than all others to lack food and money for rent, possibly because White women were more likely to have a household income over US $50,000, ensuring the ability to sustain food and housing expenses. White women were also more likely than API and AIAN women to be unable to go to the doctor or have a telemedicine appointment and report homeschooling children. Collectively, plausible explanations are that White women had competing responsibilities due to living with other family members and having a higher median number of household residents, which prevented in-person appointments with their doctor. Additionally, White women may have had privacy concerns with engaging in a phone or video telemedicine visit with their provider due to more family members or household residents being present.

Although significant differences by race/ethnicity were not identified by summed scores for the Fear of COVID-19 Scale, there were significant differences in many of the individual items, showing dimensionality within experiences of fear of COVID-19. White women consistently agreed with statements regarding psychological experiences of fear (being “very afraid” or “nervous or anxious”) relating to COVID-19, while API, Latinx, and AIAN women consistently endorsed agreement with physical manifestations of fear (“hands become clammy,” “cannot sleep,” “heart races”). Black women did not fall into these patterns, endorsing most agreement with being very afraid and experiencing anxiety relating to media about COVID-19. This may reflect differential conceptualization of fear and stress, including antidisclosure socialization and stigma around mental health [25-27] or increased symptomatology associated with adverse mental health resulting from cumulative or toxic stress [11,14].

Overall, we found high prevalence of implementation of CDC-recommended prevention behaviors. Prevalence of staying home except for essential activities was 88.6% (419/473) compared to a rate of 77.3% among US adults surveyed in May 2020 [17]. Likewise, 79.7% (377/473) of women in our study reported using a face mask, slightly higher than 74.1% of US adults surveyed in May [17]. Previous findings demonstrated decreased prevalence in physical distancing behavior from 87% in April to 79.5% in May [16,17]; in our study, 86.5% (409/473) of women reported this behavior. These differences may reflect increased uptake of prevention behaviors among women, compared to a sample of both men and women; this is consistent with findings that female gender is associated with adherence to prevention behaviors [16].

Self-isolation at home if sick was more likely among API women, possibly due to customary traditions of preventing illness. Handwashing with soap was more likely among White and API women than AIAN women. This may be due to White and API women’s increased access to education on preventing COVID-19, primarily disseminated through television and the internet. Of further importance is the possibility of performative practice of prevention behaviors among API women, in response to COVID-19–related xenophobia; individuals of Asian descent have been the targets of hate speech and crime based on the origination of COVID-19 in China [28]. Among AIAN populations, access to the internet is limited [29], and this group is further disadvantaged due to low rates of health literacy [30], limiting knowledge of prevention behaviors. Native patient navigators, similar to community health workers [31], are an avenue for education on COVID-19 prevention in AIAN communities.

API women were more likely than Black and Latinx women to not leave home for any activities, possibly due to API increased partner cohabitation, allowing them to remain indoors and rely on a partner for essential activities. Consistent with existing data, Black and Latinx populations are overrepresented in essential service industries, limiting the ability to shelter at home [10,12]. Because White women had more household residents compared to AIAN women and less partner cohabitation compared to API women, it is plausible that White women were less likely to stay home, due to the need for essential items such as food and medicine.

Supplemental analyses to assess confounding confirmed statistically significant variation between racial/ethnic groups in binary logistic regressions after adjustment for age, education, income, and type of residential community, in comparison to White participants. All adjusted odds ratios were in the expected directions and consistent with bivariate findings. However, the variation in relationship between race/ethnicity and COVID-19–related outcomes, stressors, and prevention behaviors indicate a need for further research to assess the nuanced experiences of women of all racial and ethnic identities, which was beyond the scope of this analysis, to ensure culturally responsive and appropriate public health programming.

Limitations and Strengths

The primary limitation of our research was the use of nonprobability sampling methods with recruitment through online social media, resulting in a lack of generalizability of our study findings. Thus, our sample is not representative of adult women in all US states. Our sample was younger to middle-aged, primarily resided in urban environments, and did not consist of racial/ethnic groups and other demographics such as age, education, and income that were proportionate to each US state population’s demographics. Specifically, 17.6% (82/466) of our sample resided in a rural environment, slightly underrepresenting the 19.3% of the US population in rural areas [32]. Although ethnically diverse, our sample underrepresents Latinx women while approximating the US population of Black women. Yet, API and AIAN women are overrepresented in our sample, which may add to the utility of these findings, given the underrepresentation of these groups in research. Finally,

XSL•FO
RenderX

only women with access to the internet were able to complete the survey. However, 75% of US adult women use Facebook, with three-quarters accessing Facebook daily [33], indicating that the sampling frame encompassed a significant portion of the US population.

Limitations notwithstanding, our data highlight racial/ethnic differences in COVID-19–related outcomes among adult US women. Our ability to recruit an ethnically diverse sample of women that comprised 41 US states and Washington, DC using online recruitment methods within a short time frame demonstrates the willingness of women to share their experiences on COVID-19. Future studies are needed that use probability sampling methods and longitudinal cohort study designs to examine trends over time in racial/ethnic disparities of COVID-19 outcomes and elucidate underlying factors fueling these disparities among a representative sample of adult US women.

Conclusions

The lower knowledge of testing availability, actualized testing, and prevention behaviors among Black, Latinx, and AIAN women compared to their White and API counterparts will serve to exacerbate the high rates of COVID-19 mortality already demonstrated among these populations. Meanwhile, stressors, physical manifestations of fear, and adverse mental health associated with loss of family and friends to COVID-19 may compound existing toxic stress, culminating in increasingly negative chronic health outcomes. It is vital that public health interventions for COVID-19 focus on these communities, ensuring testing, health care, and support services—including mental health, economic, and social services—are accessible and acceptable, rather than allowing for disparities to persist and worsen.

Acknowledgments

We thank the study participants who generously provided the data for this research. All authors acknowledge support from NIH/NIAID grant P30 AI036214 and the University of California, San Diego Office of Research Affairs; the Office for Equity, Diversity, and Inclusion; and the Office of the Chancellor. JKS and KMA are also supported by NIH/NIDA grant R01 AI128803. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Authors' Contributions

JKS and KMA were involved in the conceptualization, planning, design, and implementation of the study. JKS acquired funding for the study, with assistance from KMA. JKS led the drafting of the manuscript with contributions from KMA and BAW. BAW conducted the data analysis and drafted the tables. All authors reviewed and edited the manuscript for intellectual input. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics by racial/ethnic group among adult women in the United States (N=473).
[PDF File (Adobe PDF File), 236 KB - jmir_v23i7e26296_app1.pdf]

Multimedia Appendix 2
COVID-19 outcomes by racial/ethnic group among adult women in the United States (N=473).
[PDF File (Adobe PDF File), 182 KB - jmir_v23i7e26296_app2.pdf]

Multimedia Appendix 3
COVID-19 stressors and fear by racial/ethnic group among adult women in the United States (N=473).
[PDF File (Adobe PDF File), 186 KB - jmir_v23i7e26296_app3.pdf]

Multimedia Appendix 4
COVID-19 public health prevention behaviors by racial/ethnic group among adult women in the United States (N=473).
[PDF File (Adobe PDF File), 152 KB - jmir_v23i7e26296_app4.pdf]

Multimedia Appendix 5
Unadjusted and adjusted binary logistic regression models of COVID-19–related outcomes, stressors, and prevention behaviors and racial or ethnic group among adult women in the United States (N=473).
[PDF File (Adobe PDF File), 73 KB - jmir_v23i7e26296_app5.pdf]

XSL•FO
RenderX

References

1. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, et al. Coronavirus disease 2019 case surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):759-765. [doi: 10.15585/mmwr.mm6924e2] [Medline: 32555134]
2. Tai DBG, Shah A, Doubeni CA, Sia IG, Wieland ML. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. Clin Infect Dis 2021 Feb 16;72(4):703-706 [FREE Full text] [doi: 10.1093/cid/ciaa815] [Medline: 32562416]
3. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA 2020 Jun 23;323(24):2466-2467. [doi: 10.1001/jama.2020.8598] [Medline: 32391864]
4. Azar KMJ, Shen Z, Romanelli RJ, Lockhart SH, Smits K, Robinson S, et al. Disparities in outcomes among COVID-19 patients in a large health care system in California. Health Aff (Millwood) 2020 Jul;39(7):1253-1262. [doi: 10.1377/hlthaff.2020.00598] [Medline: 32437224]
5. Goldstein JR, Atherwood S. Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States. medRxiv. Preprint posted online on May 23, 2020. [doi: 10.1101/2020.05.21.20109116] [Medline: 32511557]
6. Rollston R, Galea S. COVID-19 and the social determinants of health. Am J Health Promot 2020 Jul;34(6):687-689. [doi: 10.1177/0890117120930536b] [Medline: 32551932]
7. Abrams EM, Szefler SJ. COVID-19 and the impact of social determinants of health. Lancet Respir Med 2020 Jul;8(7):659-661 [FREE Full text] [doi: 10.1016/S2213-2600(20)30234-4] [Medline: 32437646]
8. Laurencin CT, McClinton A. The COVID-19 pandemic: a call to action to identify and address racial and ethnic disparities. J Racial Ethn Health Disparities 2020 Jun;7(3):398-402 [FREE Full text] [doi: 10.1007/s40615-020-00756-0] [Medline: 32306369]
9. Abedi V, Olulana O, Avula V, Chaudhary D, Khan A, Shahjouei S, et al. Racial, economic and health inequality and COVID-19 infection in the United States. medRxiv. Preprint posted online on May 1, 2020. [doi: 10.1101/2020.04.26.20079756] [Medline: 32511647]
10. Poteat T, Millett GA, Nelson LE, Beyrer C. Understanding COVID-19 risks and vulnerabilities among black communities in America: the lethal force of syndemics. Ann Epidemiol 2020 Jul;47:1-3 [FREE Full text] [doi: 10.1016/j.annepidem.2020.05.004] [Medline: 32419765]
11. Ajilore O, Thames AD. The fire this time: the stress of racism, inflammation and COVID-19. Brain Behav Immun 2020 Aug;88:66-67 [FREE Full text] [doi: 10.1016/j.bbi.2020.06.003] [Medline: 32505712]
12. Kantamneni N. The impact of the COVID-19 pandemic on marginalized populations in the United States: a research agenda. J Vocat Behav 2020 Jun;119:103439 [FREE Full text] [doi: 10.1016/j.jvb.2020.103439] [Medline: 32390658]
13. Health equity considerations and racial and ethnic minority groups. Centers for Disease Control and Prevention. Atlanta, GA; 2020 Apr 21. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html [accessed 2020-06-18]
14. Fortuna LR, Tolou-Shams M, Robles-Ramamurthy B, Porche MV. Inequity and the disproportionate impact of COVID-19 on communities of color in the United States: the need for a trauma-informed social justice response. Psychol Trauma 2020 Jul;12(5):443-445. [doi: 10.1037/tra0000889] [Medline: 32478545]
15. Implementation of mitigation strategies for communities with local COVID-19 transmission. Centers for Disease Control and Prevention. Atlanta, GA; 2020 Apr 15. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/community-mitigation.html [accessed 2020-06-24]
16. Park CL, Russell BS, Fendrich M, Finkelstein-Fox L, Hutchison M, Becker J. Americans' COVID-19 stress, coping, and adherence to CDC guidelines. J Gen Intern Med 2020 Aug;35(8):2296-2303 [FREE Full text] [doi: 10.1007/s11606-020-05898-9] [Medline: 32472486]
17. Czeisler M, Tynan MA, Howard ME, Honeycutt S, Fulmer EB, Kidder DP, et al. Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance - United States, New York City, and Los Angeles, May 5-12, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):751-758. [doi: 10.15585/mmwr.mm6924e1] [Medline: 32555138]
18. Canning D, Karra M, Dayalu R, Guo M, Bloom DE. The association between age, COVID-19 symptoms, and social distancing behavior in the United States. medRxiv. Preprint posted online on April 23, 2020. [doi: 10.1101/2020.04.19.20065219] [Medline: 32511621]
19. Alsan M, Stantcheva S, Yang D, Cutler D. Disparities in coronavirus 2019 reported incidence, knowledge, and behavior among US adults. JAMA Netw Open 2020 Jun 01;3(6):e2012403 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12403] [Medline: 32556260]
20. Survey tool and guidance: rapid, simple, flexible behavioural insights on COVID-19. World Health Organization. 2020 Jul 29. URL: https://apps.who.int/iris/bitstream/handle/10665/333549/WHO-EURO-2020-696-40431-54222-eng.pdf [accessed 2020-08-29]
21. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The fear of COVID-19 scale: development and initial validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]

XSL•FO
RenderX

22. Arnett MJ, Thorpe RJ, Gaskin DJ, Bowie JV, LaVeist TA. Race, medical mistrust, and segregation in primary care as usual source of care: findings from the exploring health disparities in integrated communities study. J Urban Health 2016 Jun;93(3):456-467 [FREE Full text] [doi: 10.1007/s11524-016-0054-9] [Medline: 27193595]
23. Jaiswal J. Whose responsibility is it to dismantle medical mistrust? Future directions for researchers and health care providers. Behav Med 2019;45(2):188-196 [FREE Full text] [doi: 10.1080/08964289.2019.1630357] [Medline: 31343959]
24. Williamson LD, Smith MA, Bigman CA. Does discrimination breed mistrust? Examining the role of mediated and non-mediated discrimination experiences in medical mistrust. J Health Commun 2019;24(10):791-799. [doi: 10.1080/10810730.2019.1669742] [Medline: 31559916]
25. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
26. Fogel J, Ford DE. Stigma beliefs of Asian Americans with depression in an internet sample. Can J Psychiatry 2005 Jul;50(8):470-478. [doi: 10.1177/070674370505000807] [Medline: 16127965]
27. Nadeem E, Lange JM, Edge D, Fongwa M, Belin T, Miranda J. Does stigma keep poor young immigrant and U.S.-born Black and Latina women from seeking mental health care? Psychiatr Serv 2007 Dec;58(12):1547-1554. [doi: 10.1176/ps.2007.58.12.1547] [Medline: 18048555]
28. Covid-19 fueling Anti-Asian racism and xenophobia worldwide. Human Rights Watch. 2020 May 12. URL: https://www.hrw.org/news/2020/05/12/covid-19-fueling-anti-asian-racism-and-xenophobia-worldwide# [accessed 2020-07-02]
29. Klemko R. Coronavirus has been devastating to the Navajo Nation, and help for a complex fight has been slow. The Washington Post. 2020 May 16. URL: https://www.washingtonpost.com/national/coronavirus-navajo-nation-crisis/2020/05/11/b2a35c4e-91fe-11ea-a0bc-4e9ad4866d21_story.html [accessed 2020-06-29]
30. Bornstein SS, Mire RD, Barrett ED, Moyer DV, Cooney TG. The collision of COVID-19 and the U.S. health system. Ann Intern Med 2020 Sep 15;173(6):484-485 [FREE Full text] [doi: 10.7326/M20-1851] [Medline: 32484741]
31. Burhansstipanov L, Harjo L, Krebs LU, Marshall A, Lindstrom D. Cultural roles of native patient navigators for american Indian cancer patients. Front Oncol 2015;5:79. [doi: 10.3389/fonc.2015.00079] [Medline: 25984483]
32. One in five Americans live in rural areas. United States Census Bureau. 2017 Aug 09. URL: https://www.census.gov/library/stories/2017/08/rural-america.html [accessed 2020-06-29]
33. Laboratory-confirmed COVID-19-associated hospitalizations. COVID-NET. Atlanta, GA: Centers for Disease Control and Prevention; 2020 May 23. URL: https://gis.cdc.gov/grasp/COVIDNet/COVID19_5.html [accessed 2020-06-29]

Abbreviations

AIAN: American Indian or Alaskan Native
API: Asian, Native Hawaiian, or other Pacific Islander
CDC: Centers for Disease Control and Prevention
COPE: Capturing Women's Experiences in Outbreak and Pandemic Environments
REDCap: Research Electronic Data Capture
WHO: World Health Organization

Edited by G Eysenbach; submitted 06.12.20; peer-reviewed by L Hoyt D'Anna, Z Jin; comments to author 19.02.21; revised version received 12.03.21; accepted 02.04.21; published 12.07.21.

Please cite as:
Stockman JK, Wood BA, Anderson KM
Racial and Ethnic Differences in COVID-19 Outcomes, Stressors, Fear, and Prevention Behaviors Among US Women: Web-Based Cross-sectional Study
J Med Internet Res 2021;23(7):e26296
URL: https://www.jmir.org/2021/7/e26296
doi:10.2196/26296
PMID:33819909

©Jamila K Stockman, Brittany A Wood, Katherine M Anderson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Privacy-Oriented Technique for COVID-19 Contact Tracing (PROTECT) Using Homomorphic Encryption: Design and Development Study

Yongdae An[1,2], BSc; Seungmyung Lee[1], MS; Seungwoo Jung[1], BSc; Howard Park[1], MA; Yongsoo Song[3*], PhD; Taehoon Ko[4*], PhD

[1]Desilo Inc, Seoul, Republic of Korea

[2]Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea

[3]Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea

[4]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Taehoon Ko, PhD
Department of Medical Informatics
College of Medicine
The Catholic University of Korea
222 Banpo-daero
Seocho-gu
Seoul, 06591
Republic of Korea
Phone: 82 2 2258 7947
Fax: 82 508 906 5445
Email: thko@catholic.ac.kr

Abstract

Background: Various techniques are used to support contact tracing, which has been shown to be highly effective against the COVID-19 pandemic. To apply the technology, either quarantine authorities should provide the location history of patients with COVID-19, or all users should provide their own location history. This inevitably exposes either the patient's location history or the personal location history of other users. Thus, a privacy issue arises where the public good (via information release) comes in conflict with privacy exposure risks.

Objective: The objective of this study is to develop an effective contact tracing system that does not expose the location information of the patient with COVID-19 to other users of the system, or the location information of the users to the quarantine authorities.

Methods: We propose a new protocol called PRivacy Oriented Technique for Epidemic Contact Tracing (PROTECT) that securely shares location information of patients with users by using the Brakerski/Fan-Vercauteren homomorphic encryption scheme, along with a new, secure proximity computation method.

Results: We developed a mobile app for the end-user and a web service for the quarantine authorities by applying the proposed method, and we verified their effectiveness. The proposed app and web service compute the existence of intersections between the encrypted location history of patients with COVID-19 released by the quarantine authorities and that of the user saved on the user's local device. We also found that this contact tracing smartphone app can identify whether the user has been in contact with such patients within a reasonable time.

Conclusions: This newly developed method for contact tracing shares location information by using homomorphic encryption, without exposing the location information of patients with COVID-19 and other users. Homomorphic encryption is challenging to apply to practical issues despite its high security value. In this study, however, we have designed a system using the Brakerski/Fan-Vercauteren scheme that is applicable to a reasonable size and developed it to an operable format. The developed app and web service can help contact tracing for not only the COVID-19 pandemic but also other epidemics.

XSL•FO
RenderX

(J Med Internet Res 2021;23(7):e26371) doi:10.2196/26371

KEYWORDS

COVID-19; homomorphic encryption; privacy-preserving contact tracing; PROTECT protocol; GPS data; mobile application; web service

Introduction

Background

Since the first case of a previously unidentified coronavirus was reported in Wuhan, China, on December 8, 2019, the COVID-19 pandemic has affected the whole world. According to the World Health Organization (WHO), as of January 25, 2021, the number of COVID-19 cases worldwide was about 97 million, 2.1 million of which were reported to have been fatal [1]. As COVID-19 spreads rapidly, the public is experiencing growing anxiety and concern [2]. COVID-19 has incapacitated the existing medical system with its high communicability and fatality rate, and before vaccines became available, the only available countermeasures were traditional control measures, namely, case isolation, contact tracing and quarantine, physical distancing, disinfection, and practicing personal hygiene [3]; Even with vaccines now available, these countermeasures remain highly relevant and essential.

Therefore, it is imperative to understand the disease propagation and timing in order to take appropriate and timely measures. For example, when 97 patients with COVID-19 were confirmed at a call center in South Korea in March 2020, the Korea Centers for Disease Control and Prevention and the local government formed a joint response team and carried out an epidemiologic investigation using contact tracing [4]. At the time, the team identified and analyzed 1145 people and investigated their surroundings to prevent further disease transmission. Thanks to such active efforts relating to COVID-19 quarantine, despite the early onset of the COVID-19 pandemic, South Korea shows a significantly smaller number of infected patients and lower fatality rate than many other countries.

Prior Work

Ferretti et al [5] used a renewal equation formulation to develop a mathematical model for determining the speed and volume for effective screening and contact tracing necessary to stop the spread of epidemics and quantify other propagation routes. According to their study, if the self-isolation of an individual who has been in contact with a patient with COVID-19 is delayed by 3 days, no parameter combinations can achieve epidemic control. The study has mathematically proved that the epidemic can be far more effectively controlled when isolation is executed immediately or with a maximum delay of 1 to 2 days. Accordingly, this study explains that if a contact tracing application is used by a sufficient number of people, an epidemic can be controlled by maintaining temporary information about close proximity among individuals and notifying their recent contacts to initiate isolation.

An active measure against the COVID-19 pandemic requires, for example, telehealth screening and management and remote testing, but privacy regulations may pose barriers to such information dissemination. Accordingly, there are claims that privacy regulations should be relaxed for health information exchange in the context of the COVID-19 pandemic [6]. Despite the effectiveness of COVID-19 patient tracing and contact tracing by using digital tools, there are potential privacy leakage risks [7]. As a matter of fact, there have been privacy infringements in the name of public interest in South Korea during the early days of quarantine, when personal information such as gender, age, residence, and place of work was released altogether, leading to unwanted outing incidents [8].

To resolve such issues, applications and technologies are being developed that digitally execute contact tracing while protecting user privacy [9]. In some cases, GPS or Bluetooth information of mobile device users were collected in a centralized manner while attributing temporary identifiers [10,11]. In addition, there are distributed models that store the personal location history on the local mobile device only and compute the distance if a patient comes in close proximity [12]. Both methods, however, are effective only if a majority of users install the app and allow the transmission of one another's data, which in turn increases privacy risks [13]. There also exist cryptography solutions for privacy protection, such as the technology developed by Apple and Google, which utilizes secure multiparty computation without relying on a trusted server or sends anonymous encrypted or random messages [14]. The study by Gvili [15], however, claims that the said approach by Apple and Google may be vulnerable to several types of security attacks.

Study Aims

This study aims to propose the PRivacy Oriented Technique for Epidemic Contact Tracing (PROTECT) protocol for digital contact tracing that offers privacy protection by using homomorphic encryption. The proposed system exchanges location data in an encrypted format between the user and the quarantine authorities. By using a novel secure proximity computation technique, the PROTECT protocol makes it possible to identify whether the user has been in contact with any patient with COVID-19 by using only the encrypted location information. This method differs from the privacy protection technologies used in existing contact tracing systems in that it identifies the contacts with encrypted distances, and thus, it can identify whether the user has been in contact with patients with COVID-19 without exposing the user's location information. It can be said the proposed system uses a privacy-preserving technique of a higher order. In this paper, we first propose a new algorithm for proximity computation and the PROTECT protocol that utilizes this algorithm. Next, we introduce the quarantine app and web service that we have developed to apply the proposed PROTECT protocol to COVID-19 contact tracing and verify that the proposed protocol is practical through experimentation. Finally, we discuss the key results of the study, how it differs from previous studies, and its limitations.

XSL•FO
RenderX

Methods

Overview

The key to a privacy-preserving contact tracing system is to protect the location information of not only the patient but also the user, along with the ability to check for proximity. To achieve this, in this study, we used homomorphic encryption and proximity in a discrete grid system to develop a new, secure proximity computation method, and propose a new protocol called PROTECT that applies such a method to deliver data safely among the user, quarantine authorities, and the patient.

Secure Proximity Computation

The basic method to check for proximity is to compute the distance between two known locations, but this leads to unnecessary location-related privacy issues [16]. Zhong et al [17] proposed three protocols (Louis, Lester, and Pierre) that achieve privacy-secured proximity computation by employing additive homomorphic encryption. The secure proximity computation used in our proposed PROTECT protocol is inspired by the technique used in the Pierre protocol. The Pierre protocol maps the exact location information to the predefined grid areas and substitutes the proximity calculation problem to the calculation of whether the grids are identical or adjacent. It can help determine whether the two locations are in the same grid or in adjacent grids but does not provide information about the two locations. The PROTECT protocol utilizes homomorphic encryption in a novel way such that it does not expose any information other than proximity, yet it is able to perform a high-level computation that can be put into practice immediately.

Homomorphic Encryption and Brakerski/Fan-Vercauteren Scheme

Homomorphic encryption is a cryptosystem that supports computation on encrypted data. The result of encrypted computation is also a ciphertext whose decryption returns the same value as if the operation were performed over plain data. Homomorphic encryption has broad applications in cloud environments since it can be used to outsource storage and computation without data leakage.

In the last decade, there have been significant improvements in the efficiency of homomorphic encryption. Lattice-based schemes such as Brakerski-Gentry-Vaikuntanathan (BGV) [18], Brakerski/Fan-Vercauteren (BFV) [19,20], fast fully homomorphic encryption scheme over the torus (TFHE) [21] and Cheon-Kim-Kim-Song (CKKS) [22] currently yield the best performance in practice, but they provide different functionalities. In this work, we focus on the BFV scheme since the proximity of movement of patients with COVID-19 is calculated in the discrete grid system, which will be discussed later. In this system, the proximity is determined by the operation over integral vectors. The BFV scheme is efficient for vectorized operations over the integers, whereas the CKKS and TFHE schemes are more appropriate for approximate and Boolean computations, respectively. We provide a simplified description of BFV as follows.

The BFV scheme consists of five polynomial-time algorithms Setup, Enc, Dec, Add, and Mult. Note that we use symmetric-encryption, which is faster and has better noise growth compared to the public-key variant.

1. Setup (1^{λ}): For the security parameter λ, choose a parameter set and sample a secret key *sk*. Parameters include the dimension *n* and the plain-text modulus *p*.
2. Enc (*sk*, *m*): It takes as the input the secret key *sk* and a plain-text $m = (m_1,...,m_n) \in (\square_p)^n$, which is an *n*-dimensional vector over the finite field $\square_p$, and returns a ciphertext *c*.
3. Dec (*sk*, *c*): It decrypts the ciphertext *c* using the secret key *sk* and returns a plaintext *m*.
4. Add (*c*, *c'*): It outputs the addition of given ciphertexts.
5. Mult (*c*, *c'*): It performs the multiplication between given ciphertexts and returns the resulting ciphertext.

The BFV scheme satisfies the homomorphic property if parameters are chosen properly. In other words, if *c*, *c'* are encryptions of *m*, *m'*, then Add(*c*, *c'*) and Mult(*c*, *c'*) are encryptions of $m+m'$ and $m \odot m'$, respectively, where $m \odot m' = (m_1 m'_1,\ldots, m_n m'_n)$ denotes the Hadamard (component-wise) multiplication of two vectors. For simplicity, we will write Add(*c*, *c'*) = *c*+ *c'* and Mult(*c*, *c'*)= *c* * *c'*.

Proximity in Discrete Grid System

In this study, we converted the two location points to a hexagonal grid system and defined that any two points that belong to the same or adjacent grids are "proximate." The proximity between locations in a continuous space, for example, Euclidean space must be checked with comparison operations; such computation is expensive in a homomorphically encrypted system. The proximity in a discrete space, however, can be computed with a few equality checks, which can be efficiently calculated over encrypted data.

We choose the hexagonal grid system to transform the continuous location information into discrete grids. A hexagonal grid system allows for a simpler definition of neighborhood than triangular or square grids do, so as to reduce the computation overhead. As shown in Figure 1, to define a neighbor, it takes 3 classes in a triangular grid system and 2 classes in a square grid system, but just 1 class in a hexagonal grid system.

The transportation network company Uber Technologies Inc introduced a discrete global grid system called Hexagonal Hierarchical Spatial Index (H3) that is based on multiresolution hexagonal grids [23]. As shown in Figure 2, H3 provides the local IJ co-ordinate system for hexagons, which specifies a hexagonal area adjacent to the specified origin with i-axis and j-axis at an angle of 120°.

We denote by *H*: $\square$ (x, y) $\square$ (i, j) the transformation into the hexagonal grid system with a side length of . In other words, it returns the IJ co-ordinates of the hexagon to which an input point belongs. Some examples are shown in Figure 2.

XSL•FO
RenderX

Figure 1. Comparison of (A) triangular, (B) square, and (C) hexagonal grids.

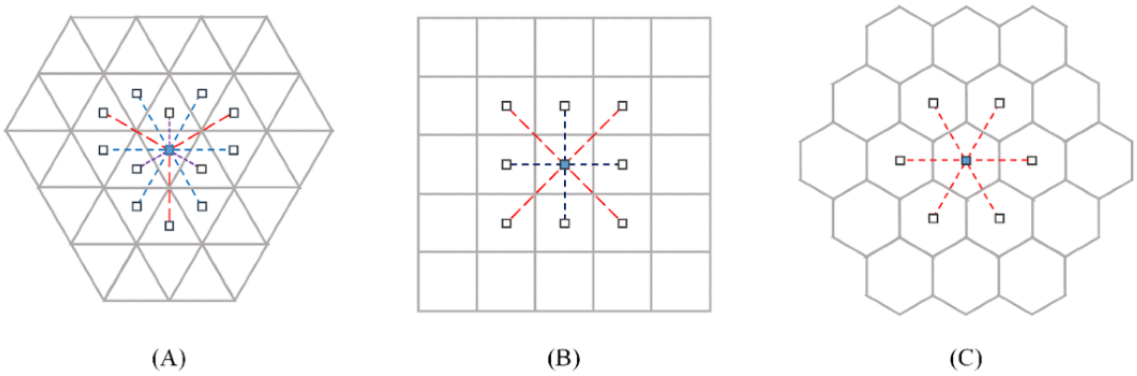
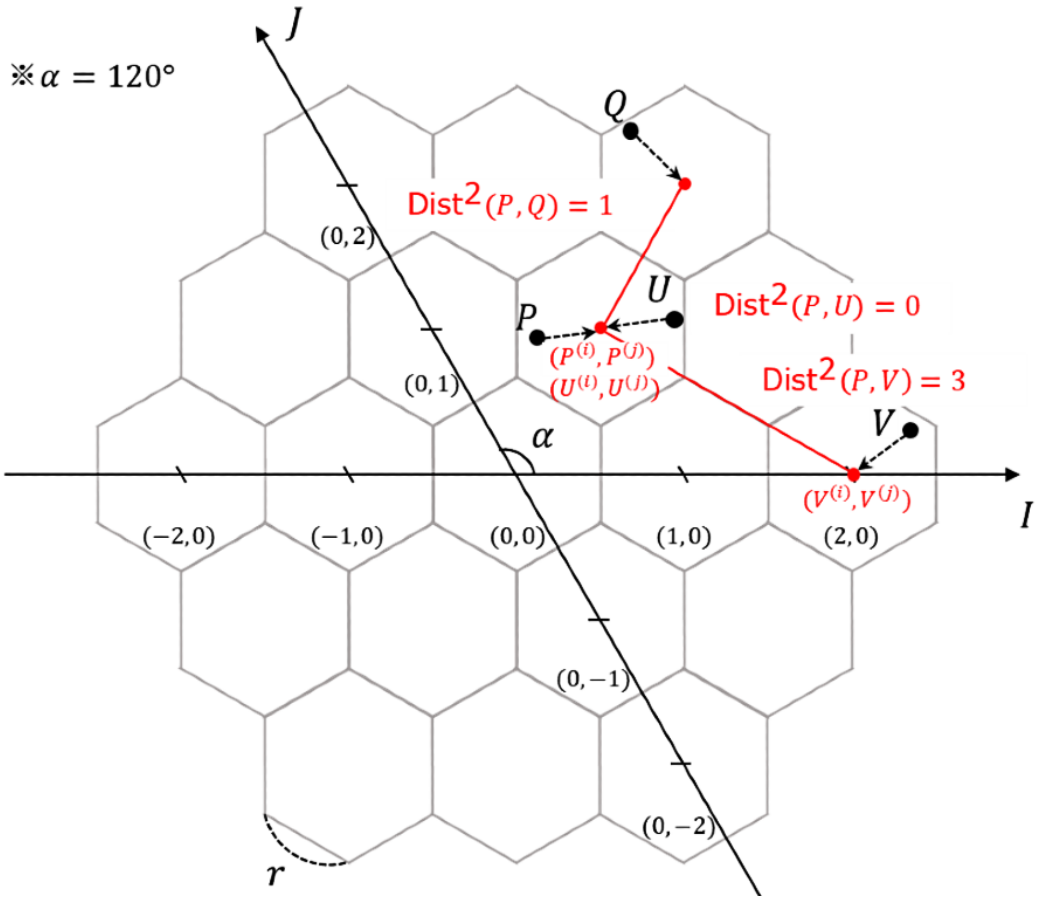


Figure 2. Local *i*, *j* coordinates of a hexagonal grid system with a side length of *r*.



We also define a metric function as follows:

$$\text{Dist}^2(\cdot,\cdot): \square, (P, Q) \square \ \| H(P) - H(Q) \|^2/(3r^2)$$

which can be computed as follows:

$$Dist^2(P,Q) = (P^{(i)} - Q^{(i)})^2 - (P^{(i)} - Q^{(i)})(P^{(j)} - Q^{(j)}) + (P^{(j)} - Q^{(j)})^2$$

where $H(P) = (P^{(i)}, P^{(j)})$, $H(Q) = (Q^{(i)}, Q^{(j)}) \in Z^2$.

We use the metric $Dist^2$ to determine the proximity between two locations. Our definition of proximity in H3 is whether the hexagonal grids corresponding to two locations and are identical or adjacent to each other, or equivalently,

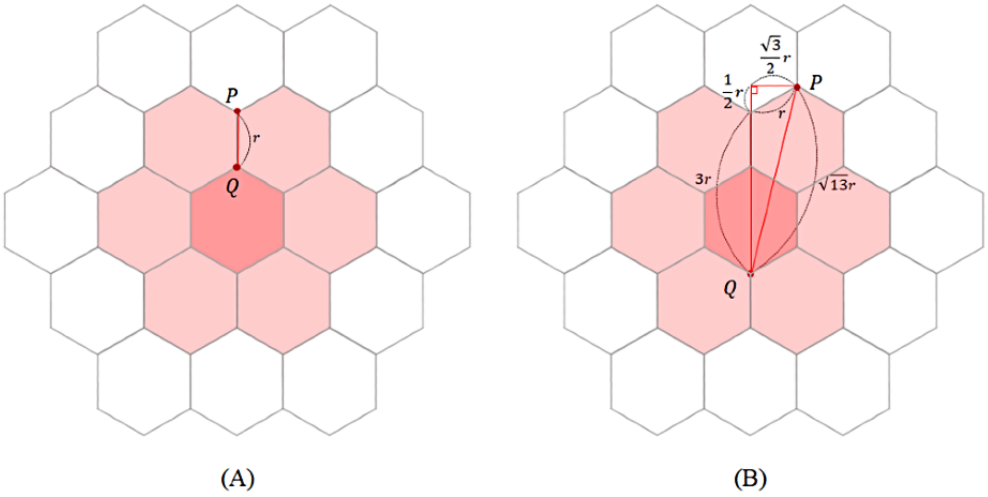
$Dist^2(P, Q) = 0$ or $Dist^2(P, Q) = 1$.

In the following, we present two properties of $Dist^2$ to convince that this is a reasonable quantity on which we can make a proper judgement.

Figure 3 shows two extreme examples where $Dist^2(P, Q)$ is relatively large/small compared to the Euclidean distance $\| P - Q \|$. In Figure 3A, the Euclidean distance between two points is but $Dist^2(P, Q) > 1$. Meanwhile, we have $\square$ and $Dist^2(P, Q) = 1$ in the case of Figure 3B. In application of contact tracing, the primary goal is to detect all contact cases, so the side length should be set sufficiently large based on if-then statements (Textbox 1).

XSL•FO
RenderX

Figure 3. Distance between two points *P* and *Q* on hexagonal grids with a side length of *r*: (A) when the distance between *P* and *Q* is slightly greater than *r* and they are not deemed proximate, (B) when the distance between *P* and *Q* is slightly less than <inline-graphic xlink:href="jmir_v23i7e26371_fig9.png" mimetype="image" xlink:type="simple"/>; and they are deemed proximate.



Textbox 1. Relationship between the approximated distance Dist^2 (P, Q) and the Euclidean distance ‖ P – Q ‖

- If $Dist^2(P, Q)) \leq 1$, then
- If ‖ P–Q ‖ < *r*, then $Dist^2(P, Q)) \leq 1$

In case of a highly contagious epidemic such as COVID-19, a single patient may cause a reproliferation; thus, the examination scope should be rather conservatively set to be broad. The WHO recommends massive testing for all suspected cases of COVID-19 [24]. The Organisation for Economic Co-operation and Development (OECD) also recommends that countries conduct as many tests for COVID-19 as possible, even if they are expensive [25]. The OECD projected that the cost of testing would be much less than the cost of a national lockdown situation [26].

PROTECT: the Proposed Protocol

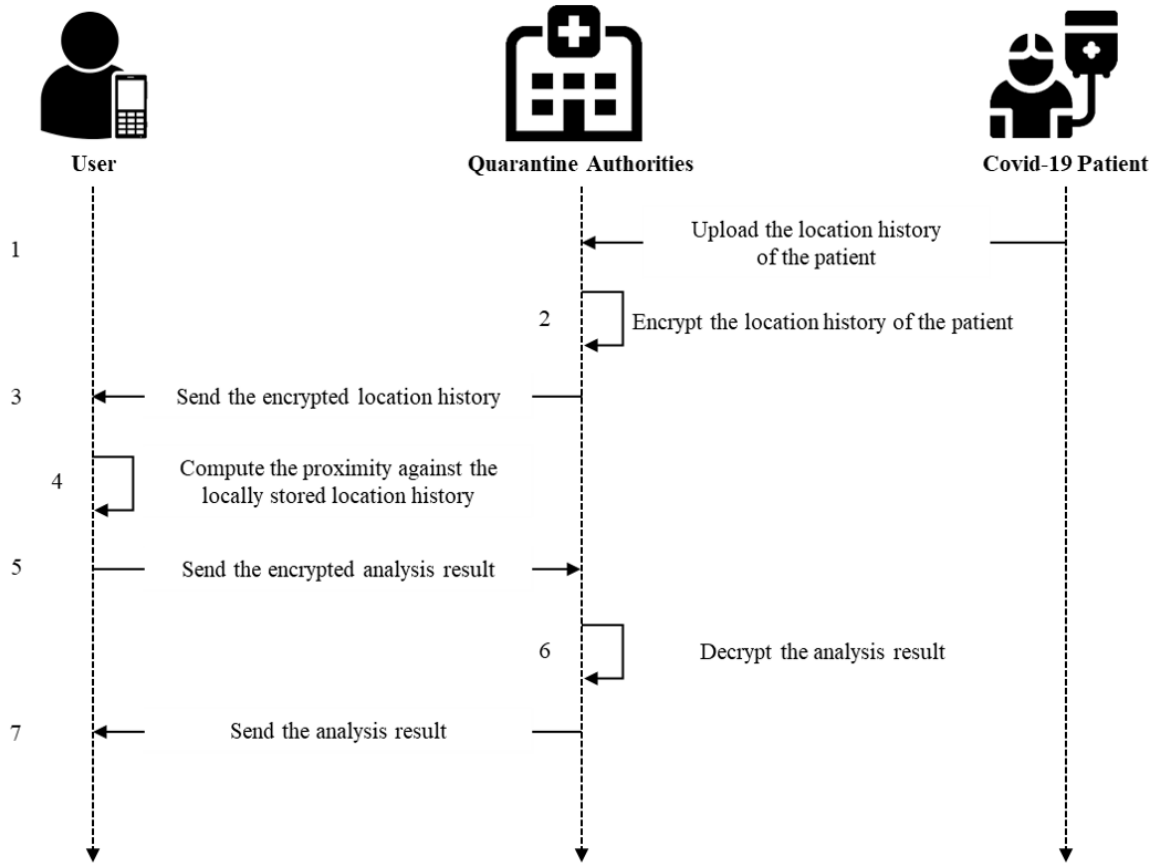
Protocol Overview

The proposed protocol PROTECT involves three parties—the user, the quarantine authorities, and the patient with COVID-19. The overall protocol flow is shown in Figure 4. The parties exchange individual ID, timestamps, and GPS locations only.

In this study, we assume that the quarantine authorities are semihonest and that the patient honestly provides their location history to the authorities. The WHO recommends that, as essential surveillance for COVID-19 considering the potential for rapid exponential growth of COVID-19 cases in populations, new cases should be identified, reported, and data included in epidemiological analysis within 24 hours. National authorities should consider including COVID-19 as a mandatory notifiable disease with requirements for immediate reporting [27]. Local governments are already collecting information to track and stop the spread of the coronavirus. The US Centers for Disease Control and Prevention has published a guideline that quarantine personnel shall investigate everyone with whom the patients with COVID-19 have had close contact during the timeframe while they may have been infectious [28], and South Korea collects the location history of patients with COVID-19 and opens them to the public so that those with a high possibility of contact with such patients can voluntarily be examined for COVID-19 [29]. Moreover, it is assumed that all communication in our protocol occurs through a secure channel. When a patient sends personal data to the authorities or the authorities send the encrypted information to the users, such exchange occurs through a secure channel that is invulnerable against a third-party attack (eg, MitM). The definition of each party and further details on the associated events are as follows.

XSL•FO
RenderX

Figure 4. Flowchart of the Privacy Oriented Technique for Epidemic Contact Tracing (PROTECT) protocol.



Patient With COVID-19

The patient is a user who has been tested positive for COVID-19 and provides 2-week GPS location history to the quarantine authorities. At this time, the location information of the patient is not encrypted.

Quarantine Authorities

The quarantine authorities are those individuals who oversee the quarantine system at the municipal or national level. The quarantine authorities receive the location information provided by the patient, encrypt it, and upload it to the server. The encrypted patient location information is then sent to the users who have installed the app. In addition, the quarantine authorities receive the result of the computation at the local user device in the encrypted format, decrypt that result, and then send the decrypted result back to the user. In this process, the quarantine authorities have no access to the personal location information stored on the local user device.

User

The user computes, on the individual local device, the proximity between their own location information and the encrypted patient location information received from the quarantine authorities. Here, homomorphic encryption makes it possible to execute computation between the encrypted location information and nonencrypted location information. The computation result is encrypted, as shown in Figure 4. The user then sends the encrypted computation result to the quarantine authorities. The user then receives and checks the decrypted computation result from the quarantine authorities and, in case of high risk of infection, is advised to follow the quarantine protocol recommended by the government.

Proximity Computation With BFV and H3

Overview

In this section, we provide technical details of proximity computation in the PROTECT protocol. Throughout this section, P_t and Q_t will denote the location data of the patient with COVID-19 and the user, respectively. The location data natively includes time information, but we suppose that data is preprocessed and synchronized so that the elements of the same index have the same timestamps.

Before the protocol starts, the quarantine authorities and the user encode their data locally into the IJ co-ordinates by using the *H*map described in the previous section and generate the vectors as follows:

$P^{(i)} = (P_1^{(i)}, \ldots, P_n^{(i)})$, $P^{(j)} = (P_1^{(j)}, \ldots, P_n^{(j)})$ and $Q^{(i)} = (Q_1^{(i)}, \ldots, Q_n^{(i)})$, $Q^{(j)} = (Q_1^{(j)}, \ldots, Q_n^{(j)})$, respectively, where $H(P_t) = (P_t^{(i)}, P_t^{(j)})$ and $H(Q_t) = (Q_t^{(i)}, Q_t^{(j)})$ for $1 \leq t \leq n$.

BFV Encryption

The server sets the parameters for BFV and generates a secret key . The server generates ciphertexts $c^{(i)} \leftarrow \mathrm{Enc}(sk, P^{(i)})$ and $c^{(j)} \leftarrow \mathrm{Enc}(sk, P^{(j)})$ by using the BFV scheme and sends them to a user.

XSL•FO
RenderX

Secure Proximity Computation

On receiving the ciphertexts, the user securely computes the proximity between and . This procedure consists of homomorphic evaluation of the proximity function followed by a ciphertext randomization process.

First, the user homomorphically evaluates $Dist^2(P, Q)$ by $c_{Dist}{}^2$ $d^{(i)} * d^{(i)} - d^{(i)} * d^{(j)} + d^{(j)} * d^{(j)} = (d^{(i)} - d^{(j)})^2 + d^{(i)} * d^{(j)}$ where $d^{(i)} = c^{(i)} - Q^{(i)}$ and $d^{(j)} = c^{(j)} - Q^{(j)}$.

This is an encryption of the vector $(Dist^2(P_t, Q_t))_{1 \le t \le n}$ from the homomorphic property of BFV. Then, the user computes and obtains $c_{Prox} = c_{Dist}{}^2 * (c_{Dist}{}^2 - 1)$. In our implementation, we performed two homomorphic multiplications after subtraction, added them, and finally performed one relinearization. Note that c_{Prox} is encrypting $Dist^2(P_t, Q_t) * (Dist^2(P_t, Q_t) - 1) = 0$ in the th slot, which is zero if and only if $Dist^2(P_t, Q_t) = 0$ or $Dist^2(P_t, Q_t) = 1$. Hence, if the user sends c_{Prox} back to quarantine authorities (the secret key owner), then they would be able to decrypt the ciphertext and determine the proximity of P_t and Q_t by checking if $Dist^2(P_t, Q_t) * (Dist^2(P_t, Q_t) - 1) = 0$ or not. However, this method is not privacy-preserving since the secret key owner can extract more information from the ciphertext c_{Prox} beyond the proximity.

Hence, the user randomizes the ciphertext c_{Prox} to solve the issue above. She generates a vector $r = (r_1, \ldots, r_n)$ whose entries r_i are sampled independently and uniformly at random from the set $\mathbb{Z}_p \setminus \{0\} = \{1, 2, \ldots, p-1\}$, and a random encryption of zero c_0 with a large noise parameter. The user outputs the ciphertext $c_{RProx} \leftarrow r * c_{Prox} + c_0$ and sends it back to the quarantine authorities. Note that the total multiplicative depth of proximity computation is 3.

Decryption

The quarantine authorities decrypt c_{RProx} and obtain $r_t' = r_t * Dist^2(P_t, Q_t) * (Dist^2(P_t, Q_t) - 1)$ for $1 \le t \le n$. They conclude that the user has been in contact with a patient at timestamp if this value is zero. We point out that the quarantine authorities learn nothing from the decrypted value about the user data more than the desired result since r_t' is purely random over $\mathbb{Z}_p \setminus \{0\}$ if $Dist^2(P_t, Q_t) \neq 0, 1$. Moreover, the ciphertext c_{RProx} itself contains no information beyond r_t' since the user randomized it by c_0. Note that the noise parameter of c_0 should be exponentially larger than that of $r * c_{Prox}$ for security proof.

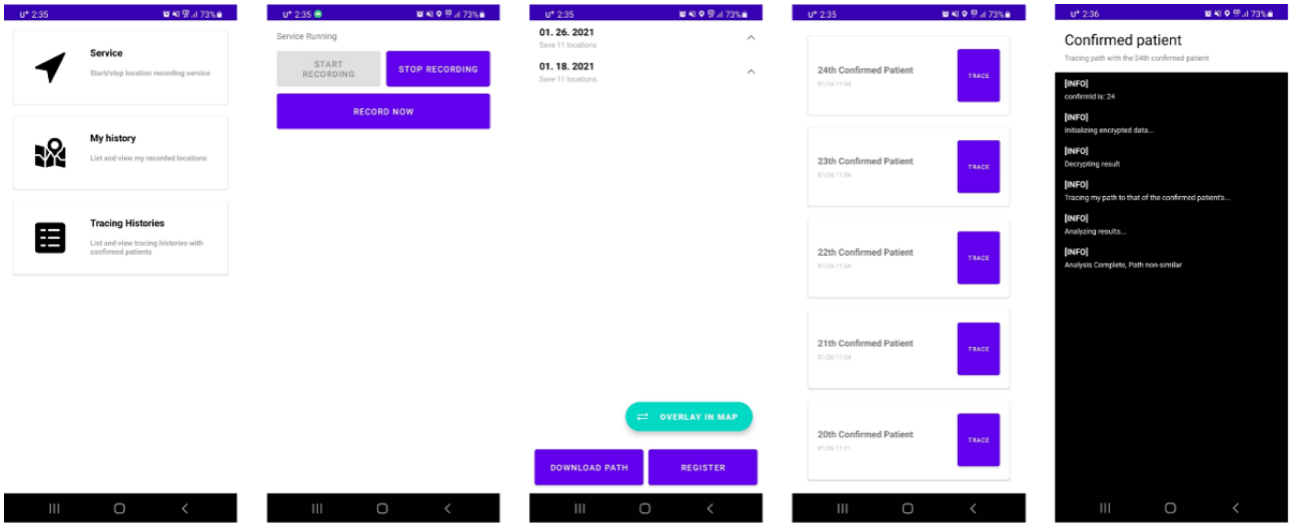
Results

In order to apply the PROTECT protocol to COVID-19 contact tracing, we have built a mobile app for patients with COVID-19 and other users, as well as a web service for the quarantine authorities. We also empirically verified the practicality of the PROTECT protocol through performance indicators related to resource consumption such as response time, CPU utilization, and memory consumption on the local device.

User App

The smartphone app for the user is as shown in Figure 5. The user can enable or disable the service any time at will (Figure 5B), and easily check the GPS information stored on the local device by date (Figure 5C). Furthermore, the user can view and compare their path with the location information of patients with COVID-19, received as a push message from the quarantine authorities, and check the location details of the potential contact in case the user is suspected to have been in contact with a patient (Figure 5D and 5E).

Figure 5. Screenshots of the user application: (A) main screen, (B) GPS data recording setup screen, (C) list of GPS data by day stored on the user's local device, (D) list of encrypted GPS data per patient received from the quarantine authorities, and (E) location history comparison and result.

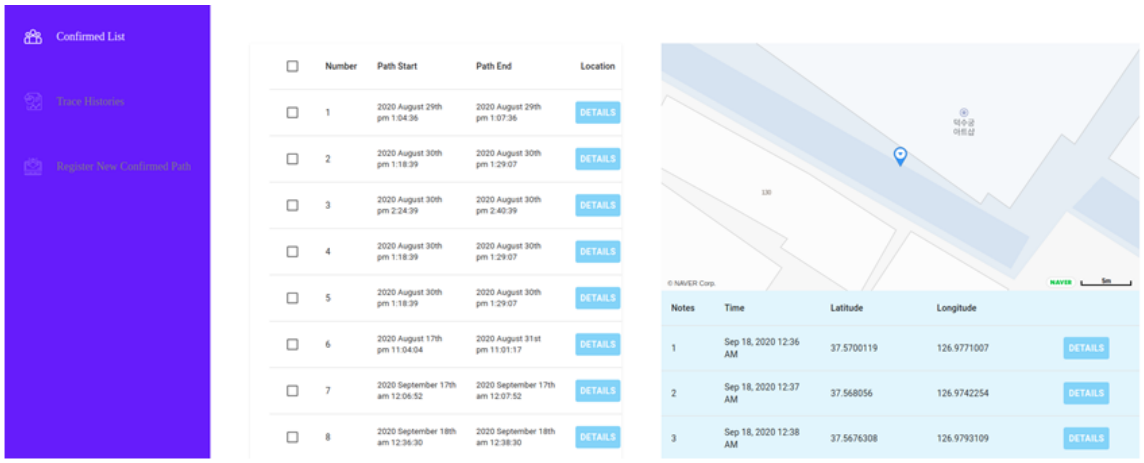


XSL•FO
RenderX

Web Service for Quarantine Authorities

The role of the quarantine authorities is to manage COVID-19 patient information and to propagate test results. For this purpose, we built a web service as shown in Figure 6. The quarantine authorities can encrypt the location information provided by the patients and propagate the encrypted information to the users who have installed the app (Figure 6A). Furthermore, although the quarantine authorities have no access to the location information of each individual user, they can analyze the results uploaded by app users and then identify the users whose location history intersects with the location history of registered patients (Figure 6B). We also developed a feature through which the quarantine authorities can easily register patient location history by manually clicking on the map in case the 2-week location data of a patient has not been recorded by the app owing to app nonuse or other reasons (Figure 6C).

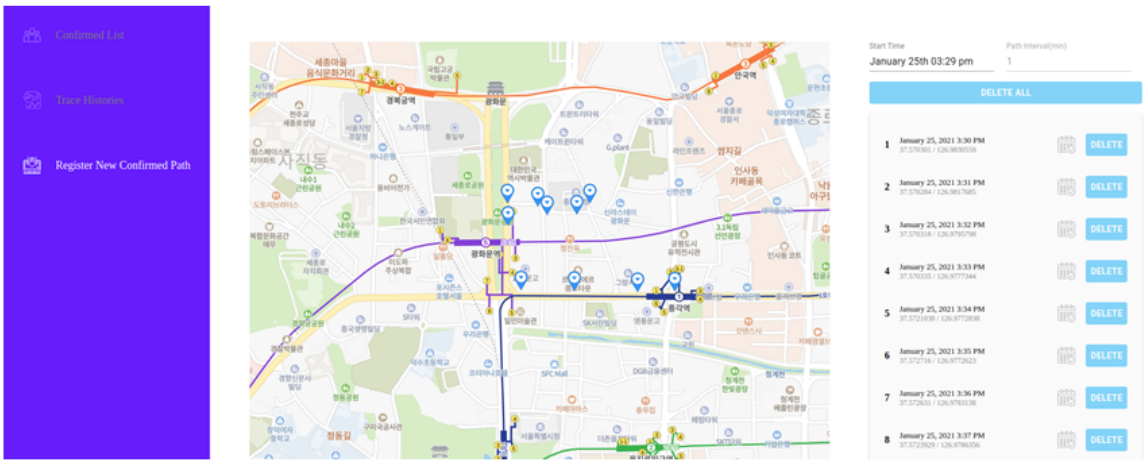
Figure 6. Screenshots of the web service developed for quarantine authorities: (A) list of confirmed patients' GPS data, (B) list of contact trace histories, and (C) register of new confirmed patients' GPS data.



(A)



(B)



(C)

XSL•FO
RenderX

Performance Indicators

To assess the practicality of the app that implements the proposed PROTECT protocol, we installed the developed app on two smartphone models—Samsung Galaxy S20 Plus and Galaxy Note 8—and conducted performance tests. The detailed specifications of the testing devices are as in Table 1.

Table 1. Specifications of testing devices.

Specifications	Galaxy S20 plus	Galaxy Note 8
Release Year	2020	2017
Chipset	Samsung Exynos 9 Octa 990	Samsung Exynos 9 Octa 8895
Processor	Octa core (2 × 2.73 GHz Mongoose + 2 × 2.5 GHz Cortex A76 + 4 × 2 GHz Cortex A55)	Octa-core (4 × 2.3 GHz Mongoose M2 + 4 × 1.7 GHz Cortex-A53)
GPU	ARM Mali-G77 MP11	ARM Mali-G71 MP20
RAM	8 GB	6 GB

To satisfy 128-bit security level while maintaining an appropriate size for computation, the base ring dimension was set to 8192, which indicates that the proximity computation for 8192 time points can be executed simultaneously. At the same time, the computation time for the entire data is determined by the size of the base ring dimension. When GPS data is collected every t seconds, the number of time points per person collected over the period of 14 days is $(60 \cdot 60 \cdot 24 \cdot 14) / t$. Surely, the larger the value of gets, the smaller the number of time points to be collected per person gets, and the number of comparison computations is also considerably reduced.

It is not necessary to use all time points to compare the time points of the user and the patient. The location information can be trimmed through various methods. It is not necessary to compare all time points for periods where the patient has stayed at a single location for a long time, such as while sleeping or working. In case many patients were present at the same location at the same time, a single computation shall suffice. Furthermore, the occasions wherein the patient has certainly made no contact, such as while driving alone, can also be excluded. Such preprocessing of data can be applied before encryption by the means of epidemiological investigation, when the quarantine authorities collect the location history data of the patients.

If the number of data points refined by the quarantine authorities is N, the number of computations (N_{comp}) is $N_{comp} = N/8192$. When the computation time for 8192 time points is $Time_{comp}$ the total time the proximity computation takes for each user (*Total_Computation_Time*) is $Total_Computation_Time = Time_{comp} * N_{comp}$.

As for the proposed PROTECT protocol, the computation times may vary depending on the processing power of the user's smart device. The test results for computation time in Samsung Galaxy S20 Plus and Note 8 are as presented in Table 2.

Table 2. Results of proximity computation tests on testing devices.

Test	Galaxy S20 Plus	Galaxy Note 8
Average CPU utilization (%)	2.158	5.425
Maximum memory consumption during computation (MB)	57.57	58.6
Computation time ($Time_{comp}$) (s)	3.246	6.967
Size of encrypted data (MB) ($TransferSize_{Q \to U}$)	1.08	1.08
Size of encrypted data (MB) ($TransferSize_{U \to Q}$)	0.814	0.814

Since S20 Plus has a more powerful processor than that of Note 8, it can be seen that is smaller. When S20 Plus is to compute 1,000,000 encrypted data points received from the quarantine authorities, $N_{comp} =$ = 123 and $Total_Computation_Time = Time_{comp} \times N_{comp} = 3.246 \times 123 = 399.258$ (seconds). Furthermore, the transfer time between users and quarantine authorities depends on the speed of the network and the size of the transferred data. To account for the difference in the network speed, we checked the size of the transferred data, the results of which are presented in Table 2. In case the quarantine authorities send 8192 encrypted patient location data points to the user, the transferred data size ($TransferSize_{Q \to U}$) is 1.08 MB on average, and when the user sends the computation result to the quarantine authorities, the data size ($TransferSize_{U \to Q}$) is 0.814 MB on average. In case of the aforementioned 1,000,000 encrypted data points, $Total_Transfer_Size_{Q \to U} = TransferSize_{Q \to U} \times N_{comp} = 1.08 \times 123 = 132.84$ (MB) and $Total_Transfer_Size_{U \to Q} = TransferSize_{U \to Q} \times N_{comp} = 0.814 \times 123 = 100.122$ (MB).

In addition, the CPU utilization level also varies depending on the processing power of the device. Samsung Galaxy S20 Plus shows a lower average CPU utilization level than that of Note

XSL•FO
RenderX

8. In case of memory consumption during computation, no significant difference was observed. As for CPU utilization and memory consumption, the proximity computation is repeated in batches of 8192, so the increase in the overall time points does not result in several fold increases.

Discussion

Principal Results

This study proposed the PROTECT protocol, which utilizes homomorphic encryption to protect privacy while performing digital contact tracing. For this, a novel secure proximity computation technique has been developed so that the location data can be encrypted and exchanged between the user and the quarantine authorities, while the potential COVID-19 patient contact can be identified with encrypted distances only. This method differs from the privacy protection measures used in existing contact tracing systems in that it identifies contacts with encrypted distances, enabling a far higher level of privacy-preserving contact tracing. Our proposed protocol assumes the existence of a centralized organization that already collects the location history of patients and checks for proximity without exposing the location information of the patient to the user or that of the user to the organization. The Bluetooth-based method proposed by Apple and Google requires adoption by a majority of the population for contact tracing to take effect. Our proposed protocol, however, can exhibit the effect of contact tracing for those who have installed the app, no matter how small the number of such users is, provided that the quarantine organization encrypts and provides the patient data collected so far. In addition, the user does not have to provide their location information to the government, which is an advantage against psychological repulsion, one of the greatest hindrances against promoting the use of such an app.

Furthermore, in order to apply the PROTECT protocol to the COVID-19 pandemic, we built a mobile app for patients and users and a web service for the quarantine authorities. In addition, the performance indicators related to resource consumption, such as computation time, CPU utilization, and memory consumption, verify that this protocol is practical enough to be applied to actual COVID-19 quarantine measures.

Comparison With Contact Tracing in Euclidean Space

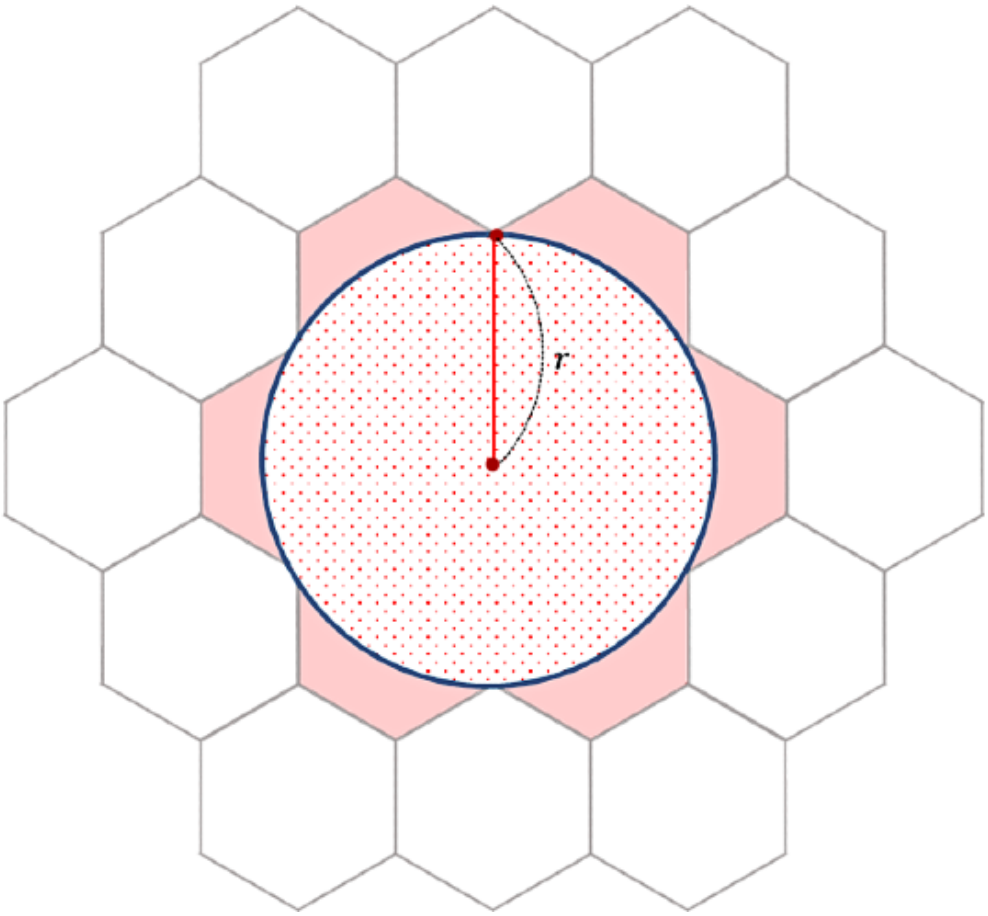
Contact tracing in Euclidean space is not secure in terms of privacy. To check for proximity under the Euclidean system, one must first compute the Euclidean distance between the two known locations. This, however, leads to an unnecessary location privacy issue. In order to calculate proximity between the locations of two parties, whoever executes that calculation—be it one of the two parties or an entirely separate third party—must possess the location information of both parties. This implies that at least one party must reveal their location information to another party. On the other hand, as previously discussed, the PROTECT protocol can only determine that two locations are in the same or adjacent grid through secure proximity computation.

Since the hexagonal grid system recognizes a wider range as adjacent than the Euclidean distance method, contact tracing in Euclidean space might appear to be more efficient than the PROTECT protocol. Suppose that we need to test everyone who is within a Euclidean distance of r or less from the location of the patient with COVID-19. As shown in Figure 7, if the side length of a single hexagonal grid is $r/2$, the area of 7 hexagonal grids is . The area of the circle is πr^2 so that the rate (deemed adjacent in the hexagonal grid but not actually adjacent in the Euclidean space) is about 30.9%. If the length of one side of the hexagonal grid is made smaller, this ratio decreases, but the secure computation time in the PROTECT protocol increases. However, the simulation was made under naive but inevitable assumptions.

As COVID-19 is highly contagious, the spread of COVID-19 cannot be covered by the Euclidean space. Rather, the examination scope should be expanded sufficiently. As mentioned earlier, many international organizations are already recommending mass testing for COVID-19 [25-27]. Previous studies have also demonstrated that mass testing is highly effective through COVID-19 epidemic simulation [30,31].

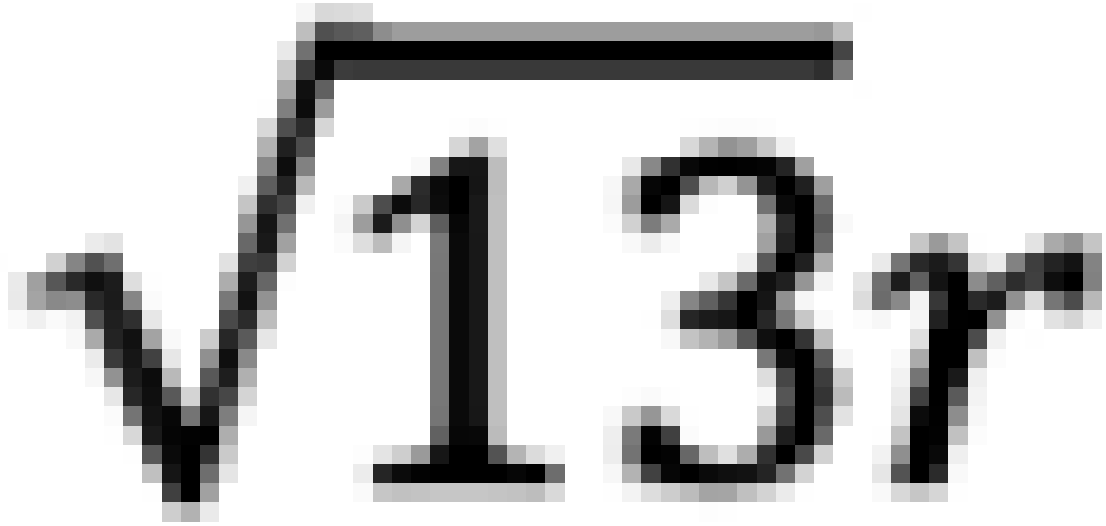
XSL•FO
RenderX

Figure 7. A circle with radius *r* and 7 hexagons with r/2 sides.



XSL•FO
RenderX

Figure 9. Inline graphic 2.



Limitations

The proposed protocol and system also have limitations shared by all contact tracing methods that make use of digital technologies. First, there is the limitation of the performance of the smartphone device itself. The accuracy of the GPS location data of each device may vary. GPS, especially, when compared to Bluetooth, is relatively less accurate in an urban setting with many indoor environments and high-rise buildings. Such limitations of the device performance can be complemented by using indoor positioning data such as Wi-Fi and Bluetooth, as well as geomagnetic location measurement techniques. In fact, the indoor positioning data collection technologies have seen much improvement through the advancement of technologies such as fingerprinting.

Second, in this study, all patients with COVID-19 are considered. However, in actual quarantine scenarios, we only need to compare to the patients in the corresponding region, thus reducing the total time taken for the comparison. Third, the homomorphically encrypted computation logic was developed in the same language for both the web server and the mobile app. Thus, there were inefficiencies to make it run within an Android app, such as porting Microsoft SEAL (Simple Encrypted Arithmetic Library) to WebAssembly with a JavaScript interface and then running it on a JavaScript engine within a browser. This should be addressed by directly invoking SEAL C++ APIs (application programming interfaces) using JNI (Java Native Interface) for Android applications. Resolving such inefficiencies would enable the development of a practical solution that is applicable to actual quarantine scenarios.

Comparison with Prior Work

In order to prevent the location privacy issue related to the calculation of proximity using location information, Gruteser and Grunwald [32] and Bettini et al [33] utilized the concept of k-anonymity for location privacy through dummy data. This method can be a useful means to protect location privacy in various location-based services. However, it is inefficient in the practical setting where the proximity needs to be checked while protecting the location privacy of both parties. In addition, Hu et al [34] proposed a method of calculating the distance using homomorphic encryption and a comparative computation using Geohash. The first method can prevent direct exposure of location information through its use of homomorphic encryption, but location information can be indirectly inferred from the distance information obtained in the end; thus, it cannot be deemed sufficiently safe in terms of location privacy. The second method compares geohash using homomorphic encryption, and it is thus relatively much safer in terms of location privacy. However, it is neither practical nor efficient in terms of computation owing to its use of bitwise homomorphic computation.

XSL•FO
RenderX

In contrast, the secure proximity computation method used in this study substitutes the problem of proximity calculation with the computation of identity or adjacency of two grids by mapping the exact location information to a predefined grid system, and then executes the calculation under homomorphic encryption, thus being safe in terms of location privacy and excellent in terms of computation.

Furthermore, from a system-wise perspective, most existing apps, such as TraceTogether [10] of Singapore or COVIDSafe [12] of Australia, are effective only if the users install the app and allow the exchange of data among one another, and they have the drawback of increased privacy risks. Moreover, the method proposed by Apple and Google is also vulnerable to several types of security attacks [15]. Above all, the previously mentioned methods become effective only when a majority of users use the app. However, with regard to the app and web service based on the proposed PROTECT protocol, even if there is only a single user, that user can effectively identify the occurrences of patient contacts as long as the central quarantine authorities have collected the patient location history.

Conclusions

The whole world is facing an unprecedented global pandemic situation and is trying to overcome this crisis by all means. Various information technology solutions are being actively suggested in this context. Owing to the potential risk of privacy leaks, however, the adoption rates are low and there has been no case of a *killer app* actively used by many.

In this study, we have described the development of a new proximity computation algorithm that can identify proximity occurrences without exposing the COVID-19 patient location and the user location to one another by homomorphically encrypting the location information. We propose PROTECT, a privacy-preserving contact tracing protocol that uses this algorithm, for use during the current COVID-19 pandemic. In order to apply this protocol to COVID-19 quarantine measures, the proposed protocol has been implemented as a smartphone app for patients and the public and a web service for quarantine authorities. Homomorphic encryption of the BFV scheme is used to design a system applicable to a reasonable scale, and through experiments under various conditions, it has been verified that this service is practical enough to be implemented in a real-world scenario. We hope that this approach that intends to resolve the issue through new technologies contributes to the early discovery and suppression of other potential diseases in future.

Conflicts of Interest

None declared.

References

1. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-01-25]
2. Lee H, Noh EB, Choi SH, Zhao B, Nam EW. Determining public opinion of the COVID-19 pandemic in South Korea and Japan: social network mining on Twitter. Healthc Inform Res 2020 Oct;26(4):335-343 [FREE Full text] [doi: 10.4258/hir.2020.26.4.335] [Medline: 33190468]
3. Bae YS, Kim KH, Choi SW, Ko T, Jeong CW, Cho B, et al. Information technology-based management of clinically healthy COVID-19 patients: lessons from a living and treatment support center operated by Seoul National University Hospital. J Med Internet Res 2020 Jun 12;22(6):e19938 [FREE Full text] [doi: 10.2196/19938] [Medline: 32490843]
4. Park SY, Kim Y, Yi S, Lee S, Na B, Kim CB, et al. Coronavirus disease outbreak in call center, South Korea. Emerg Infect Dis 2020 Aug;26(8):1666-1670 [FREE Full text] [doi: 10.3201/eid2608.201274] [Medline: 32324530]
5. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936. [doi: 10.1126/science.abb6936] [Medline: 32234805]
6. Lenert L, McSwain B. Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jun 01;27(6):963-966 [FREE Full text] [doi: 10.1093/jamia/ocaa039] [Medline: 32232432]
7. Park S, Choi GJ, Ko H. Information technology-based tracing strategy in response to COVID-19 in South Korea-privacy controversies. JAMA 2020 Jun 02;323(21):2129-2130. [doi: 10.1001/jama.2020.6602] [Medline: 32324202]
8. Jo EA. South Korea's experiment in pandemic surveillance. The Diplomat. 2020 Apr 13. URL: https://thediplomat.com/2020/04/south-koreas-experiment-in-pandemic-surveillance/ [accessed 2021-01-25]
9. Wang D, Liu F. Privacy risk and preservation for COVID-19 contact tracing apps. arXiv. Preprint posted online on June 27, 2020 [FREE Full text]
10. TraceTogether – A Singapore Government Agency website. URL: https://www.tracetogether.gov.sg/ [accessed 2021-05-20]
11. Bay J, Kek J, Tan A, Hau C, Yongquan L, Tan J. BlueTrace: a privacy-preserving protocol for community-driven contact tracing across borders. BlueTrace Protocol. 2020. URL: https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf [accessed 2021-01-25]
12. COVIDSafe app - Department of Health, Australian Government. URL: https://www.health.gov.au/resources/apps-and-tools/covidsafe-app [accessed 2021-01-25]

XSL•FO
RenderX

13. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: privacy considerations and related trade-offs. arXiv. Preprint posted online on March 30, 2020 [FREE Full text]
14. Privacy-Preserving Contact Tracing: Apple and Google. URL: https://covid19.apple.com/contacttracing/ [accessed 2021-01-25]
15. Gvili Y. Security analysis of the COVID-19 contact tracing specifications by Apple Inc. and Google Inc. Cryptology ePrint Archive, Report 2020/428. 2020 Sep 26. URL: https://eprint.iacr.org/2020/428 [accessed 2021-01-25]
16. Beresford A, Stajano F. Location privacy in pervasive computing. IEEE Pervasive Comput 2003 Jan;2(1):46-55. [doi: 10.1109/mprv.2003.1186725]
17. Zhong G, Goldberg I, Hengartner U. Louis, Lester and Pierre: three protocols for location privacy. In: Privacy Enhancing Technologies. PET 2007. Lecture Notes in Computer Science, vol 4776. Berlin, Heidelberg: Springer; 2007 Presented at: International Workshop on Privacy Enhancing Technologies. Springer, Berlin, Heidelberg June; 2007; Ottawa, ON p. 62-76. [doi: 10.1007/978-3-540-75551-7_5]
18. Brakerski Z, Gentry C, Vaikuntanathan V. (Leveled) Fully homomorphic encryption without bootstrapping. ACM Trans Comput Theory 2014 Jul;6(3):1-36. [doi: 10.1145/2633600]
19. Fan J, Frederik V. Somewhat practical fully homomorphic encryption. Cryptology ePrint Archive, Report 2012/144. 2012 Mar 17. URL: https://eprint.iacr.org/2012/144 [accessed 2021-05-19]
20. Brakerski Z. Fully homomorphic encryption without modulus switching from classical GapSVP. In: Safavi-Naini R, Canetti R, editors. Advances in Cryptology – CRYPTO 2012. Lecture Notes in Computer Science, vol 7417. Berlin, Heidelberg: Springer; 2012.
21. Chillotti I, Gama N, Georgieva M, Izabachène M. Faster fully homomorphic encryption: bootstrapping in less than 0.1 seconds. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016 Nov 09 Presented at: International Conference on the Theory and Application of Cryptology and Information Security. Springer, Berlin, Heidelberg December; December 4, 2016; Hanoi, Vietnam p. 3-33. [doi: 10.1007/978-3-662-53887-6_1]
22. Cheon J, Kim A, Kim M, Song Y. Homomorphic encryption for arithmetic of approximate numbers. Cham: Springer International Publishing; 2017 Nov 30 Presented at: International Conference on the Theory and Application of Cryptology and Information Security. Springer, Cham December; December 3, 2017; Hong Kong, China p. 409-437. [doi: 10.1007/978-3-319-70694-8_15]
23. Brodsky I. H3: Uber's hexagonal hierarchical spatial index. Uber Engineering. 2018 Jun 27. URL: https://eng.uber.com/h3/ [accessed 2020-11-27]
24. Laboratory testing strategy recommendations for COVID-19: interim guidance. World Health Organization. 2020 Mar 21. URL: https://www.who.int/publications/i/item/laboratory-testing-strategy-recommendations-for-covid-19-interim-guidance [accessed 2021-01-25]
25. Allain-Dupré D, Chatry I, Michalun V, Moisio A. The territorial impact of COVID-19: managing the crisis across levels of government. In: OECD Policy Responses to Coronavirus. Paris: OECD Publishing; Nov 10, 2020.
26. Scarpetta S, Pearson M, Colombo F, Guanais F. Testing for COVID-19: a way to lift confinement restrictions. In: OECD Policy Responses to Coronavirus. Paris: OECD Publishing; May 04, 2020.
27. Surveillance strategies for COVID-19 human infection. Interim guidance. World Health Organization. 2020 Dec 16. URL: https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2020.8 [accessed 2021-01-25]
28. Case investigation and contact tracing: part of a multipronged approach to fight the COVID-19 pandemic. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/principles-contact-tracing.html [accessed 2021-03-05]
29. Jung G, Lee H, Kim A, Lee U. Too much information: assessing privacy risks of contact trace data disclosure on people with COVID-19 in South Korea. Front Public Health 2020 Jun 18;8:305 [FREE Full text] [doi: 10.3389/fpubh.2020.00305] [Medline: 32626681]
30. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
31. Alvarez MM, González-González E, Trujillo-de Santiago G. Modeling COVID-19 epidemics in an Excel spreadsheet to enable first-hand accurate predictions of the pandemic evolution in urban areas. Sci Rep 2021 Feb 22;11(1):4327 [FREE Full text] [doi: 10.1038/s41598-021-83697-w] [Medline: 33619337]
32. Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking. 2003 May 05 Presented at: The 1st International Conference on Mobile Systems, Applications and Services; May 5, 2003; San Francisco, CA p. 31-42. [doi: 10.1145/1066116.1189037]
33. Bettini C, Wang XS, Jajodia S. Protecting privacy against location-based personal identification. 2005 Sep 02 Presented at: The Second VDLB International Conference on Secure Data Management; 2005-09-02; Trondheim, Norway p. 185-199. [doi: 10.1007/11552338_13]
34. Hu P, Mukherjee T, Valliappan A, Radziszowski S. Homomorphic proximity computation in geosocial networks. 2016 Apr 10 Presented at: IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE April; October 4, 2016; San Francisco, CA p. 616-621. [doi: 10.1109/infcomw.2016.7562150]

XSL•FO
RenderX

Abbreviations

API: application programming interface
BFV: Brakerski/Fan-Vercauteren
BGV: Brakerski-Gentry-Vaikuntanathan
CKKS: Cheon-Kim-Kim-Song
H3: Hexagonal Hierarchical Spatial Index
JNI: Java Native Interface
OECD: Organisation for Economic Co-operation and Development
PROTECT: Privacy Oriented Technique for Epidemic Contact Tracing
SEAL: Simple Encrypted Arithmetic Library
TFHE: fast fully homomorphic encryption scheme over the torus
WHO: World Health Organization

Edited by C Basch; submitted 30.01.21; peer-reviewed by J Li, J Ropero; comments to author 24.02.21; revised version received 17.03.21; accepted 29.04.21; published 12.07.21.

Please cite as:
An Y, Lee S, Jung S, Park H, Song Y, Ko T
Privacy-Oriented Technique for COVID-19 Contact Tracing (PROTECT) Using Homomorphic Encryption: Design and Development Study
J Med Internet Res 2021;23(7):e26371
URL: https://www.jmir.org/2021/7/e26371
doi:10.2196/26371
PMID:33999829

©Yongdae An, Seungmyung Lee, Seungwoo Jung, Howard Park, Yongsoo Song, Taehoon Ko. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study

Colton Margus[1,2], MD; Natasha Brown[1,2], MD; Attila J Hertelendy[1,3], PhD; Michelle R Safferman[4,5], MD; Alexander Hart[1,2], MD; Gregory R Ciottone[1,2], MD

[1]Division of Disaster Medicine, Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[2]Department of Emergency Medicine, Harvard Medical School, Boston, MA, United States

[3]Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

[4]Department of Emergency Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]Department of Emergency Medicine, Mount Sinai Morningside-West, New York, NY, United States

Corresponding Author:
Colton Margus, MD
Division of Disaster Medicine
Department of Emergency Medicine
Beth Israel Deaconess Medical Center
Rosenberg Bldg. 2nd Floor
One Deaconess Road
Boston, MA, 02215
United States
Phone: 1 6177543462
Email: cmargus@bidmc.harvard.edu

Related Article:

Comment in: https://www.jmir.org/2022/3/e34870

Abstract

Background: The early conversations on social media by emergency physicians offer a window into the ongoing response to the COVID-19 pandemic.

Objective: This retrospective observational study of emergency physician Twitter use details how the health care crisis has influenced emergency physician discourse online and how this discourse may have use as a harbinger of ensuing surge.

Methods: Followers of the three main emergency physician professional organizations were identified using Twitter's application programming interface. They and their followers were included in the study if they identified explicitly as US-based emergency physicians. Statuses, or tweets, were obtained between January 4, 2020, when the new disease was first reported, and December 14, 2020, when vaccination first began. Original tweets underwent sentiment analysis using the previously validated Valence Aware Dictionary and Sentiment Reasoner (VADER) tool as well as topic modeling using latent Dirichlet allocation unsupervised machine learning. Sentiment and topic trends were then correlated with daily change in new COVID-19 cases and inpatient bed utilization.

Results: A total of 3463 emergency physicians produced 334,747 unique English-language tweets during the study period. Out of 3463 participants, 910 (26.3%) stated that they were in training, and 466 of 902 (51.7%) participants who provided their gender identified as men. Overall tweet volume went from a pre-March 2020 mean of 481.9 (SD 72.7) daily tweets to a mean of 1065.5 (SD 257.3) daily tweets thereafter. Parameter and topic number tuning led to 20 tweet topics, with a topic coherence of 0.49. Except for a week in June and 4 days in November, discourse was dominated by the health care system (45,570/334,747, 13.6%). Discussion of pandemic response, epidemiology, and clinical care were jointly found to moderately correlate with COVID-19 hospital bed utilization (Pearson r=0.41), as was the occurrence of "covid," "coronavirus," or "pandemic" in tweet texts (r=0.47). Momentum in COVID-19 tweets, as demonstrated by a sustained crossing of 7- and 28-day moving averages, was found to have

XSL•FO
RenderX

occurred on an average of 45.0 (SD 12.7) days before peak COVID-19 hospital bed utilization across the country and in the four most contributory states.

Conclusions: COVID-19 Twitter discussion among emergency physicians correlates with and may precede the rising of hospital burden. This study, therefore, begins to depict the extent to which the ongoing pandemic has affected the field of emergency medicine discourse online and suggests a potential avenue for understanding predictors of surge.

(J Med Internet Res 2021;23(7):e28615) doi:10.2196/28615

KEYWORDS

COVID-19 pandemic; emergency medicine; disaster medicine; crisis standards of care; latent Dirichlet allocation; topic modeling; Twitter; sentiment analysis; surge capacity; physician wellness; social media; internet; infodemiology; COVID-19

Introduction

The contagiousness, fatality rate, and long-term sequelae thus far attributed to COVID-19, the disease caused by SARS-CoV-2, have led to significant strains on the health care system. Since the World Health Organization (WHO) first reported "a cluster of pneumonia cases" in Wuhan, China, on January 4, 2020 [1], the social media platform Twitter has become a source of both official health information and unofficial medical discourse regarding the ongoing pandemic. Boasting 180 million daily active users [2], not only does the service allow account holders to share links, media, and brief strings of text, but it has evolved into a public forum for unvetted information that can augment, if not supersede, more traditional dissemination methods.

On December 11, 2020, the US Food and Drug Administration (FDA) used Twitter to announce its authorization for immediate emergency use of the COVID-19 vaccine developed by Pfizer and BioNTech [3]. In its Twitter message, or *tweet*, about the decision, the FDA (@US_FDA) reiterated its aim to "assure the public and medical community that it has conducted a thorough evaluation of the available safety, effectiveness, and manufacturing quality information" [4]. Directly addressing Twitter's medical community in this way was intentional: throughout the COVID-19 pandemic, many physicians turned to social media rather than traditional medical information channels to discuss the merits and demerits of possible treatments, prior to the availability of formal clinical guidance. Myriad treatment modalities and prevention strategies have been proposed at all levels, and Twitter has served as a means of disseminating everything from guidelines and data to anecdotes and opinions [5].

Utilizing social media to aid in the mapping of an ongoing crisis is not new, and Twitter use has previously been linked to, among others, the H1N1 and Zika virus epidemics [6,7]. Yet even with unparalleled international effort, formal forecasting models of COVID-19 have largely failed [8], and many geopolitical comparisons in popular media now in hindsight appear to have been premature [9-12]. As the front and, for many Americans, only door into the US health care system, emergency departments continue to be looked to for public health surveillance and treatment strategies, as a kind of finger on the epidemiological pulse of their communities [13,14].

Emergency physicians, in particular, have long been at the forefront of physician engagement with social media, relying on a budding network of fellow clinicians collaborating on what has become known as free open-access medical education [15]. The COVID-19 pandemic only further accentuates the unique role of the emergency physician community online, as frontline providers who not only take on substantial risk but who may also be able to provide substantial insight. Facing changed admission criteria, expanded alternate care sites, and recycled equipment, emergency physicians have been forced into the unenviable position of making difficult triaging and resource allocation decisions. This study, therefore, seeks to characterize the sentiment and topic trends in emergency physician discourse on Twitter throughout the prevaccination pandemic, as a potential harbinger of the surge needs that followed.

Methods

Sampling and Data Collection

This work was approved as exempt human subjects research through the Beth Israel Deaconess Medical Center Institutional Review Board in Boston, Massachusetts. In order to access Twitter's application programming interface (API), a developer account was applied for and obtained. Python 3.8.5 and the Tweepy library (Python Software Foundation) [16] then made it possible to acquire all unique followers of the three major physician professional societies in emergency medicine: the American College of Emergency Physicians (ACEP; @ACEPNow), the Society for Academic Emergency Medicine (SAEM; @SAEMonline), and the American Academy of Emergency Medicine (AAEM; @aaeminfo) [17-21]. Because sex and gender are not directly recorded by Twitter but have previously been shown to influence social media engagement and even clinical diagnosis and management [22-24], gendered nouns and pronouns stated in user bios were considered in their place. Those users with privacy settings that would render tweets protected from analysis were removed. Each user bio was then initially screened by textual search for including any of the 157 text strings decided by the research team as connoting a public acknowledgement of one's role as an emergency physician, such as "emergency medicine physician," "emergency D.O.," or "ER doc."

Exclusion criteria included aspiring emergency physicians and students, organizations, physicians from other specialties, as well as users belonging to other professions, living outside the United States, or without a clear location at the state level. However, emergency physicians still in training, whether described as interns or residents, were not excluded. Exclusion for any of these reasons was determined by two practicing

XSL•FO
RenderX

emergency physicians each reviewing and sorting all users manually and independently, with any discrepancies decided by consensus.

A chain-referral sampling method was then employed in order to expand the study group to include those US-based emergency physicians on Twitter not following one of the major professional organization accounts [25]. Followers of already-included participants were then aggregated to create a composite list of potential additional participants. After applying the same exclusions, this new group of users was then appended to the original as a more comprehensive sampling of US emergency physicians on Twitter.

All available tweets up to Twitter's own limit of 3200 per user were acquired for each study participant. Tweets were removed if reposted as a *retweet* from a different post, if non-English, or if falling outside the study period from and including January 4, 2020, based on the date of the initial WHO announcement, to and including December 14, 2020, based on the date of the first FDA-approved vaccination in the United States [26].

Sentiment and Topic Generation

Several different methods have previously been employed to conduct sentiment analysis of tweets specific to the health care field, with 46% of such tweets demonstrating sentiment of some kind [27]. Here, the open-source Valence Aware Diction and Sentiment Reasoner (VADER) analysis tool was used to determine both the direction and extent of tweet sentiment polarity, based on a lexicon of sentiment-related words. VADER has been shown to outperform human raters and, in handling emoji and slang, is particularly suited for social media text [28,29]. Sentiment polarity ratings were summed and standardized as a compound score between –1 and 1. By convention, tweets with a compound score between –0.05 and 0.05 were classified as neutral [29,30].

Using the gensim Python package, all tweets were tokenized and preprocessed, including removal of punctuation, special characters, mentions of other users, *stop words* of little topic value, and links to external websites. Hashtags, which users sometimes use to denote a contextual theme [31,32], were converted to text. Because frequently co-occurring words can exist with unique meaning, two- or three-word phrases were also considered as independent tokens, as in "healthcare_workers." These preprocessed tweets then underwent unsupervised topic modeling in order to discern meaningful content themes. Latent Dirichlet allocation (LDA) is a common method for topic modeling that has previously been utilized to analyze health care–related tweets [33,34]. Topic coherence, as proposed by Röder et al due to its higher correlation with human topic ranking [35], was then maximized by iteratively modeling over a range of topic numbers as well as α parameters. The resulting topic model was then used to assign a dominant topic to each tweet included in the sample.

In order to contextualize these sentiment and topic trends within the ongoing pandemic, daily COVID-19 case counts and COVID-19 inpatient bed utilization (CIBU) rates were acquired from the US Centers for Disease Control and Prevention and the US Department of Health and Human Services [36,37]. These data were converted to 7-day simple moving averages to account for lower weekend reporting and other daily fluctuations [38,39]. Tweet volume, sentiment, and dominant topic trends were then correlated with new COVID-19 cases and CIBU through Prism, version 9.0.2 (GraphPad Software).

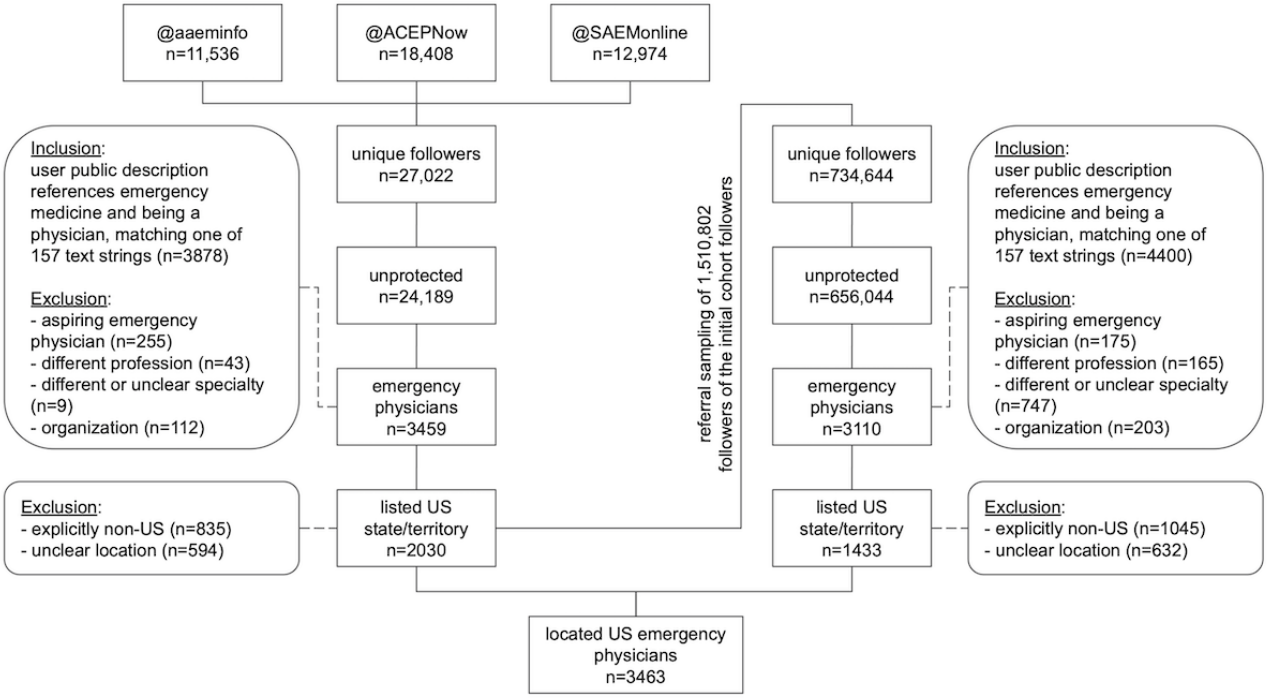
Further comparison between public health and Twitter data was made possible by plotting the 7- and 28-day simple moving averages and observing their intersection as a potential indicator for momentum using Excel, version 16.47.1 (Microsoft). Similarly, a moving average convergence/divergence oscillator (MACD) was generated by subtracting the 28-day exponential moving average from the 7-day exponential moving average. This MACD was then monitored for both (1) turning positive and (2) crossing above its own 7-day exponential moving average. These cross signals based on simple moving averages and on the MACD are both loosely derived from lagging indicators of historical price patterns that are commonly used in finance to guide investment decisions and have previously been applied directly to SARS-CoV-2 infection data [40,41].

Results

The three key US emergency physician organizations had a collective 42,918 followers of their primary Twitter accounts as of December 11, 2020. When those following more than one professional organization were only counted once, there were 27,022 unique followers, with 10,905 (40.4%) belonging to at least two of the three groups (Figure 1). As an approximation for cohesion, the overlap coefficient of the three handles was 0.43, calculated as the ratio of the intersection over the maximum possible intersection ($[A \cap B \cap C]/\min[|A|,|B|,|C|]$) [42]. After exclusions, 2073 US-based emergency physicians were identified, with high interrater reliability (κ=0.96).

XSL•FO
RenderX

Figure 1. Overview of the methodology applied for study participant selection. Unprotected unique followers of the Twitter handles for three key US emergency physician professional organizations were sampled; they were included if referencing being an emergency medicine physician and excluded if not found to be an individual emergency physician located in a US state or territory. A referral sample of the original sample's followers underwent the same inclusion and exclusion criteria to contribute additional US-based emergency physicians to the study group. AAEM: American Academy of Emergency Medicine; ACEP: American College of Emergency Physicians; SAEM: Society for Academic Emergency Medicine.



There were 1,510,802 followers of the initial cohort acquired on December 12 and 13, 2020, with 734,644 found to be internally unique as well as distinct from the original user list assessed. Applying the same inclusion and exclusion criteria resulted in 3110 emergency physicians, 1433 of whom could clearly be identified as located in specific US states, territories, or districts through their public Twitter location and description (κ=0.94). Combining the two groups, there were 3463 US-based emergency physicians included in the study.

Study participants had been using Twitter for an average of 6.6 (SD 3.5) years, with an average of 183.8 (SD 491.0) total tweets (Table 1). Only 910 out of 3463 (26.3%) participants explicitly described themselves as a resident or intern currently in training. The most common US states represented were New York (433/3463, 12.5%) and California (395/3463, 11.4%), and the most contributory US region was the Northeast (1057/3463, 30.5%). Self-identified gender was infrequent (902/3463, 26.0%), with 466 of those 902 participants (51.7%) identifying as a man.

XSL•FO
RenderX

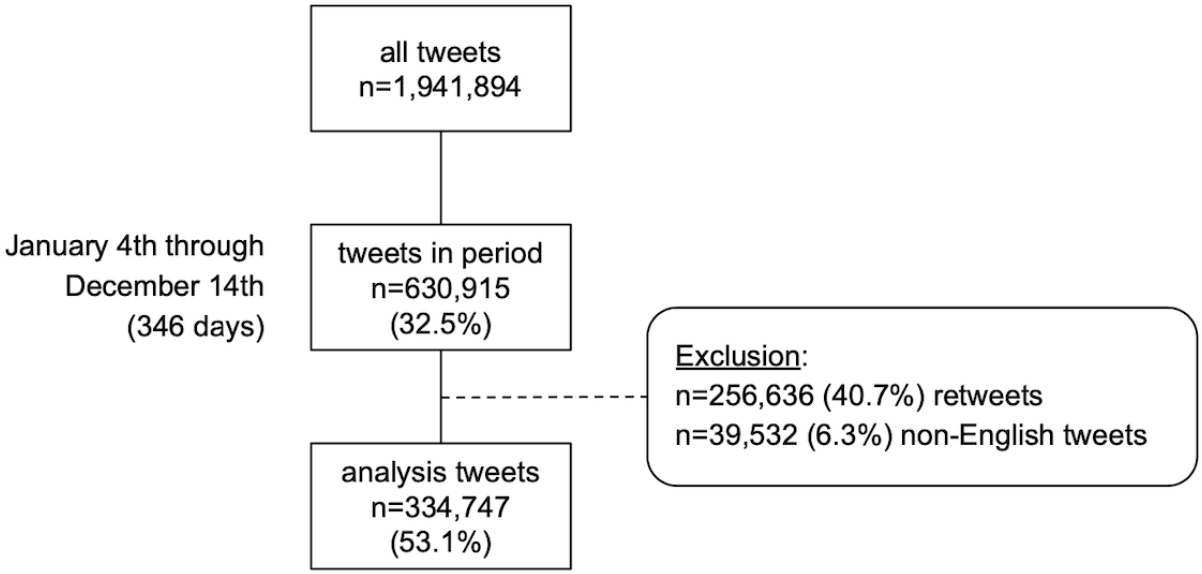
Table 1. Descriptive statistics of included US-based emergency physicians on Twitter.

Characteristic	Value (N=3463)
Gender, n (%)	
Identified	902 (26.0)
Men (n=902)	466 (51.7)
Women (n=902)	436 (48.3)
Unidentified	2561 (74.0)
Usage	
Verified account, n (%)	27 (0.8)
Duration (years), mean (SD)	6.6 (3.5)
Tweets, mean (SD)	183.8 (491.0)
Followers, mean (SD)	664.6 (5326.3)
Since 2007-2009, n (%)	519 (15.0)
Since 2010-2014, n (%)	1471 (42.5)
Since 2015-2019, n (%)	1235 (35.7)
Since 2020, n (%)	238 (6.9)
Organizations followed, n (%)	
American Academy of Emergency Medicine (AAEM) only	144 (4.2)
American College of Emergency Physicians (ACEP) only	351 (10.1)
Society of Academic Emergency Medicine (SAEM) only	275 (7.9)
AAEM and ACEP	114 (3.3)
AAEM and SAEM	148 (4.3)
ACEP and SAEM	343 (9.9)
All three organizations	655 (18.9)
None	1433 (41.4)
Training: identified as in training, n (%)	910 (26.3)
US region, n (%)	
Midwest	789 (22.8)
Northeast	1057 (30.5)
South	884 (25.5)
West	724 (20.9)
Territory	9 (0.3)
Top five US states, n (%)	
New York	433 (12.5)
California	395 (11.4)
Pennsylvania	249 (7.2)
Texas	235 (6.8)
Illinois	212 (6.1)

Tweets collected for the study group totaled 1,941,894 as of December 24, 2020, with 630,915 (32.5%) of those obtained falling between January 4 and December 14, inclusive (Figure 2). Because of a cap on the number of tweets able to be pulled through the official Twitter API, 44 out of 3463 (1.3%) users appeared to have exceeded the limit, such that not all tweets would have been captured. Despite truncation, these avid users still contributed 140,938 of all 1,941,894 (7.3%) collected tweets. Overall, 256,636 (40.7%) retweets and 39,532 (6.3%) non-English tweets were removed, leaving 334,747 (53.1%) unique English-language tweets for analysis. Daily volume went from a pre-March mean of 481.9 (SD 72.7) tweets to a mean of 1065.5 (SD 257.3) tweets thereafter.

XSL•FO
RenderX

Figure 2. Overview of the methodology applied for tweets selected for analysis. Tweets collected for study participants were included if they fell within the January 4 through December 14, 2020, study period and excluded if they were found to be retweets, non-English tweets, and tweets of indeterminate language.



After preprocessing, 1,958,230 semantic units, or *tokens*, were found within the corpus of tweets, with a total vocabulary of 12,401. LDA modeling over a range of topic numbers and parameters settled on a total of 20 content topics for this study. Two physicians then worked together to manually and jointly label these 20 topics based on discussion of key terms and representative tweets. For example, the topic with the top five terms of "resident," "student," "residency," "learn," and "year" was labeled as *medical training*. In this way, the most prevalent topics were found to relate to the *health care system* (45,570/334,747, 13.6%), *collaboration* (20,112/334,747, 6.0%), and *politics* (18,186/334,747, 5.4%) (Table 2). Notably, the *health care system* was the dominant topic throughout the study period, with two exceptions: it was supplanted from June 2 to 9 by *race relations* and from November 6 to 10 by *politics*.

Daily change in 7-day moving averages for specific tweet topics and sentiment polarity demonstrated small Pearson correlation coefficients for the topics of pandemic response (r=0.26, 95% CI 0.15-0.36), epidemiology (r=0.25, 95% CI 0.14-0.35), and clinical care (r=0.23, 95% CI 0.13-0.33) with reported COVID-19 cases (all P<.001) (Table 2). There was greater correlation for these topics with hospital bed utilization (all P<.001). While the three topics considered jointly were even more correlated (r=0.41, 95% CI 0.31-0.50), they still fell short of the correlation seen with the 9.4% of included tweets containing "covid," "coronavirus," "corona virus," "cov-2," "cov2," or "pandemic" within the tweet text (r=0.47, 95% CI 0.38-0.56) (all P<.001). A proportional stacked area chart reveals an early overall increase in Twitter use as daily case counts rose, particularly among COVID-19–related topics (Figure 3). Aggregated sentiment scores reached a nadir on June 6, when *race relations* was the dominant topic in the sample, and again on October 6, the day after then–US President Donald Trump was discharged from his COVID-19 hospital admission [43].

XSL•FO
RenderX

Table 2. Topic descriptive statistics.

Topic label	Total tweets (N=334,747), n (%)	Compound sentiment score, mean (SD)	Case Pearson correlation, *r* (95% CI)	CIBU[a] Pearson correlation, *r* (95% CI)	CIBU Spearman correlation, *r* (95% CI)	Key terms	Example tweet
Health care system	45,570 (13.6)	0.16 (0.36)	0.03 (–0.07 to 0.14)	0.12 (0.00 to 0.23)	0.07 (–0.05 to 0.18)	Care, health, physician, medicine, practice, system, medical, important, community, change, issue, lead, work, support, research, address, focus, policy, create, improve	"I'm not the one to ask about nursing. Nursing has always defined itself. The problem is the definition used to define 'advanced nursing' is the same definition used to define medicine. That is not the same definition that was used years ago, it changed. Common sense dictates one"
Collaboration	20,112 (6.0)	0.58 (0.34)	0.01 (–0.10 to 0.12)	0.09 (–0.03 to 0.20)	0.11 (–0.01 to 0.23)	Work, great, amazing, team, love, proud, congrat, congratulation, colleague, job, awesome, friend, support, part, good, hard, share, incredible, today, honor	"Honored to receive this award from @TXChildrensPEM[b] section. Thank you all for being such a great group of mentors, colleagues, and friends! Also, winning the Fellow's Award means so much. Happy for such a great group of fellows and mentees!"
Pandemic care	13,240 (4.0)	0.14 (0.50)	0.23 (0.13 to 0.33)	0.26 (0.15 to 0.37)	0.18 (0.06 to 0.29)	Patient, care, hospital, covid, doctor, nurse, emergency, doc, physician, call, sick, staff, visit, admit, treat, ed[c], icu[d], medical, work, room	"Physician-owned hospitals can increase the number of licensed beds, operating rooms, and procedure rooms by converting observation beds to inpatient beds, among other means, to accommodate patient surge"
Research	16,415 (4.9)	0.02 (0.47)	0.07 (–0.04 to 0.17)	0.06 (–0.06 to 0.17)	0.07 (–0.05 to 0.19)	Patient, study, treatment, high, give, low, risk, drug, pain, show, dose, trial, present, treat, disease, early, diagnosis, benefit, med, effect	"Take-homes from 2020 ACEP[e] Opioids Clinical Policy: 1. Treat opioid withdrawal with buprenorphine. 2. Preferentially prescribe non-opioids for acute pain. 3. Avoid prescribing opioids for chronic pain. 4. Do not prescribe sedatives to patients taking opioids"
Race relations	15,128 (4.5)	–0.17 (0.51)	0.00 (–0.10 to 0.12)	0.03 (–0.09 to 0.14)	–0.02 (–0.14 to 0.09)	People, black, man, kill, woman, call, speak, matter, police, stand, white, stop, racism, word, racist, history, protest, happen, die, wrong	"Black lives matter means Blackqueerlives matter, Black trans lives matter, Black non-binary lives matter, Black femmelives matter, Black incarcerated lives matter, and Black disabled lives matter..."
Pandemic response	14,143 (4.2)	0.05 (0.50)	0.26 (0.15 to 0.36)	0.38 (0.28 to 0.47)	0.27 (0.16 to 0.38)	Covid, pandemic, coronavirus, vaccine, response, protect, health, virus, fight, ppe[f], crisis, die, continue, country, worker, spread, leadership, expert, action, state	"#COVID. COVID COVID COVID COVID COVID COVID COVID COVID 183,000+ Americans dead, and counting... Care for your neighbors. #WearAMask"

XSL•FO
RenderX

Topic label	Total tweets (N=334,747), n (%)	Compound sentiment score, mean (SD)	Case Pearson correlation, *r* (95% CI)	CIBU[a] Pearson correlation, *r* (95% CI)	CIBU Spearman correlation, *r* (95% CI)	Key terms	Example tweet
Reading	17,897 (5.3)	0.27 (0.41)	0.06 (–0.05 to 0.17)	0.20 (0.09 to 0.31)	0.13 (0.01 to 0.25)	Read, great, check, write, thread, article, post, book, list, follow, find, share, good, send, add, paper, twitter, email, tweet, link	“Please read the first paragraph of the new image again. It literally is saying what I originally replied with. Google searches do no good if you won’t read the text of what you find, not just the header.”
Schedule	15,577 (4.7)	0.15 (0.41)	0.04 (–0.07 to 0.15)	0.10 (–0.02 to 0.21)	0.07 (–0.05 to 0.19)	Day, time, week, today, hour, start, work, shift, year, long, wait, month, back, night, spend, end, run, sleep, minute, morning	“The length of shifts of studies in this paper started at 13 hours. Time off during day hours not post-night is obviously not the same as working 13 hours and having a few hours off before bed.”
Public safety	11,594 (3.5)	0.19 (0.45)	0.05 (–0.06 to 0.16)	0.26 (0.15 to 0.36)	0.12 (0.00 to 0.24)	People, school, safe, open, home, work, close, place, stay, follow, mask, risk, live, order, family, plan, community, back, person, kid	“Every single store we went into on Michigan Ave required a mask. Our hotel requires a mask anywhere inside. Even Millenium Park requires a mask to enter and walk around outside. And on the streets plenty of people are masked outside. I think compliance is excellent”
Politics	18,186 (5.4)	–0.01 (0.49)	0.00 (–0.11 to 0.11)	–0.01 (–0.13 to 0.10)	–0.05 (–0.17 to 0.07)	Vote, trump, election, lie, country, state, people, lose, president, win, debate, biden, stop, count, call, support, political, campaign, american, fact	“Trump’s personal lawyer: Guilty. Trump's campaign manager: Guilty. Trump’s deputy campaign manager: Guilty. Trump’s National Security Advisor: Guilty. Trump’s political advisor: Guilty.”
Entertainment	18,100 (5.4)	0.17 (0.47)	0.03 (–0.07 to 0.14)	0.05 (–0.07 to 0.16)	0.09 (–0.03 to 0.20)	Watch, good, play, love, guy, game, thing, time, bad, video, pretty, show, give, favorite, big, real, season, fan, idea, listen	“I only watched pro sports and news for decades, never watching any of the popular TV shows; now I’ve actually started watching Downton Abbey instead. I guess Breaking Bad or GOT is next. I haven’t seen a single episode of either. Any other suggestions?”
Epidemiology	14,208 (4.2)	0.01 (0.48)	0.25 (0.14 to 0.35)	0.36 (0.25 to 0.45)	0.17 (0.05 to 0.28)	Covid, test, case, death, number, testing, people, high, positive, report, rate, day, virus, infection, risk, coronavirus, symptom, spread, increase, rise	“Q: what if I traveled to high risk area/ contact w known #COVID19 case) & HAVE symptoms? A: Isolate yourself. U meet testing criteria but do not HAVE to get tested. If u test negative for everything, please isolate yourself until symptoms resolve as for any contagious illness.”

XSL•FO
RenderX

Topic label	Total tweets (N=334,747), n (%)	Compound sentiment score, mean (SD)	Case Pearson correlation, *r* (95% CI)	CIBU[a] Pearson correlation, *r* (95% CI)	CIBU Spearman correlation, *r* (95% CI)	Key terms	Example tweet
Scientific inquiry	13,235 (4.0)	0.13 (0.46)	–0.04 (–0.15 to 0.07)	0.12 (0.00 to 0.23)	0.12 (0.01 to 0.24)	Question, agree, datum, point, answer, science, study, base, evidence, fact, show, understand, opinion, true, wrong, important, good, correct, information, clear	"Many, including @realDonaldTrump, have abandoned science, logic and common sense Don't take medical advice from charlatans Listen to real experts Hydroxychloroquine data shows no benefit + significant potential harms"
Protective equipment	15,156 (4.5)	0.14 (0.42)	0.07 (–0.04 to 0.18)	0.15 (0.03 to 0.26)	0.14 (0.03 to 0.26)	Mask, wear, put, hand, face, line, eye, time, head, find, leave, back, hold, room, pull, run, clean, cover, hair, remove	"A woman on the subway just pulled her mask down to blow her nose. Feeling like somehow people still don't get it..."
Business of medicine	12,217 (3.6)	0.12 (0.48)	0.07 (–0.04 to 0.18)	0.06 (–0.06 to 0.17)	0.12 (0.00 to 0.24)	Pay, money, system, physician, cost, make, free, health, state, give, problem, care, medical, job, insurance, company, hospital, healthcare, cut, plan	"Benchmarking to INW[g] rates or lower, based on a antiquated federal fee scheduling system, is a nonstarter for most physician owned and operated practices. Incentivize competition in the marketplace. Offer better reimbursement rates than CMGs[h] or large groups. Break monopolies."
Family	14,296 (4.3)	0.20 (0.46)	0.12 (0.01 to 0.23)	0.12 (0.00 to 0.23)	0.15 (0.03 to 0.26)	Year, kid, child, friend, family, good, call, give, time, talk, feel, parent, young, today, make, back, mom, remember, wife, baby	"Same with my wife and her parents back in the day.younger sister got everything she wanted. We married young and never asked for anything. Only her mother came to our wedding (teen marriage never lasts) 44 years ago...no wedding gifts."
Lifestyle	17,610 (5.3)	0.18 (0.42)	–0.02 (–0.13 to 0.09)	0.07 (–0.05 to 0.18)	0.07 (–0.05 to 0.19)	Make, eat, food, car, good, run, water, dog, drive, walk, buy, love, drink, bring, hot, coffee, nice, thing, cool, enjoy	"Stuffed peppers: Cut 4 bell peppers in half lengthwise. In a skillet saute 2 cups spinach, 1/3 white onion and garlic. Add 1lb ground chicken. Season to taste. Add 1 cup cauliflower rice. Stuff the 'rice' into the peppers. Top peppers w/ cheese & bake for 20mins on 375 degrees."
Medical training	15,994 (4.8)	0.36 (0.42)	0.08 (–0.03 to 0.19)	0.13 (0.01 to 0.24)	–0.03 (–0.15 to 0.09)	Resident, student, residency, learn, year, program, medical, medtwitter, great, join, today, virtual, mede, attend, interview, teach, school, talk, match, conference	"Thankful for my residency family today! Had a great week of shifts and an awesome virtual conference last week! My faculty and coresidents have been so amazing these last few months!"

XSL•FO
RenderX

Topic label	Total tweets (N=334,747), n (%)	Compound sentiment score, mean (SD)	Case Pearson correlation, *r* (95% CI)	CIBU[a] Pearson correlation, *r* (95% CI)	CIBU Spearman correlation, *r* (95% CI)	Key terms	Example tweet
Emotional reaction	12,401 (3.7)	0.14 (0.49)	–0.03 (–0.14 to 0.08)	0.10 (–0.01 to 0.22)	0.11 (–0.01 to 0.23)	Make, thing, good, feel, time, bad, people, hard, lot, happen, agree, change, easy, hear, decision, part, point, find, real, sense	“Are you nervous? Lots of people feel nervous when they come here That’s normal What are you nervous about? Are you nervous that something may hurt? A lot of people worry about that Nothing is going to hurt right now If that changes I’ll tell you & we’ll get thru it”
Inspirational	13,668 (4.1)	0.30 (0.50)	0.05 (–0.06 to 0.16)	0.14 (0.03 to 0.25)	0.17 (0.05 to 0.28)	Life, love, feel, hope, true, live, world, word, human, story, time, share, save, real, experience, moment, heart, family, find, change	“Thought of the day: I can share my earthly riches like peace, joy, time, talents, giftings, physical helps, hope, wisdom, emotional strength, encouragement, etc.”

[a]CIBU: COVID-19 inpatient bed utilization.

[b]TXChildrensPEM: Texas Children’s Hospital Pediatric Emergency Medicine.

[c]ed: emergency department.

[d]icu: intensive care unit.

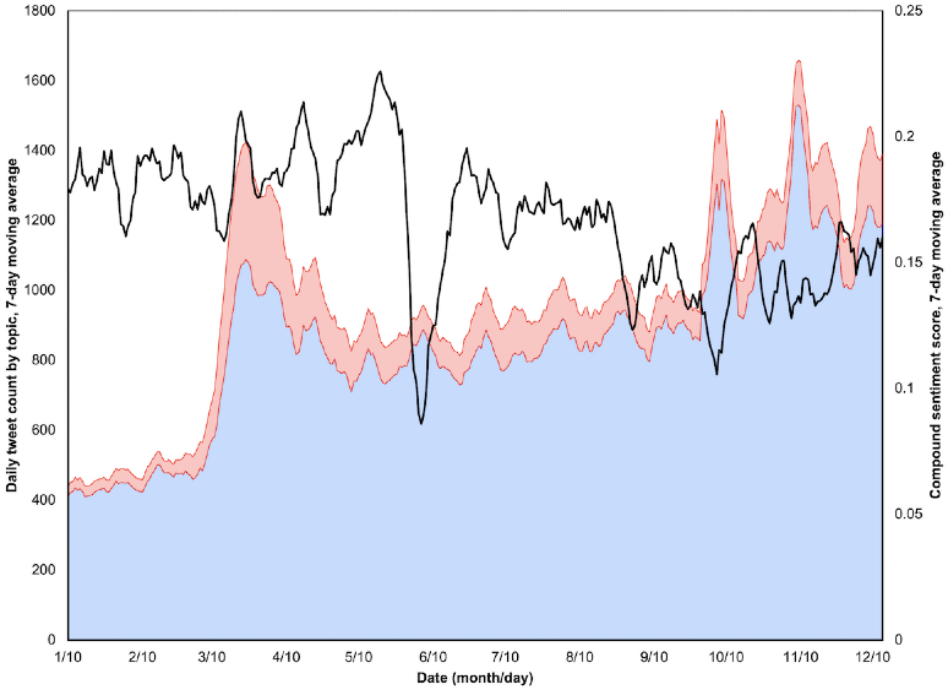
[e]ACEP: American College of Emergency Physicians.

[f]ppe: personal protective equipment.

[g]INW: in-network.

[h]CMG: contract management group.

Figure 3. Stacked area plot of 7-day moving average daily counts of latent Dirichlet allocation–derived topics, both those pertaining to COVID-19 (red area) and those not (blue area) (left axis), plotted against the 7-day moving average of daily compound sentiment scores nationally (right axis).

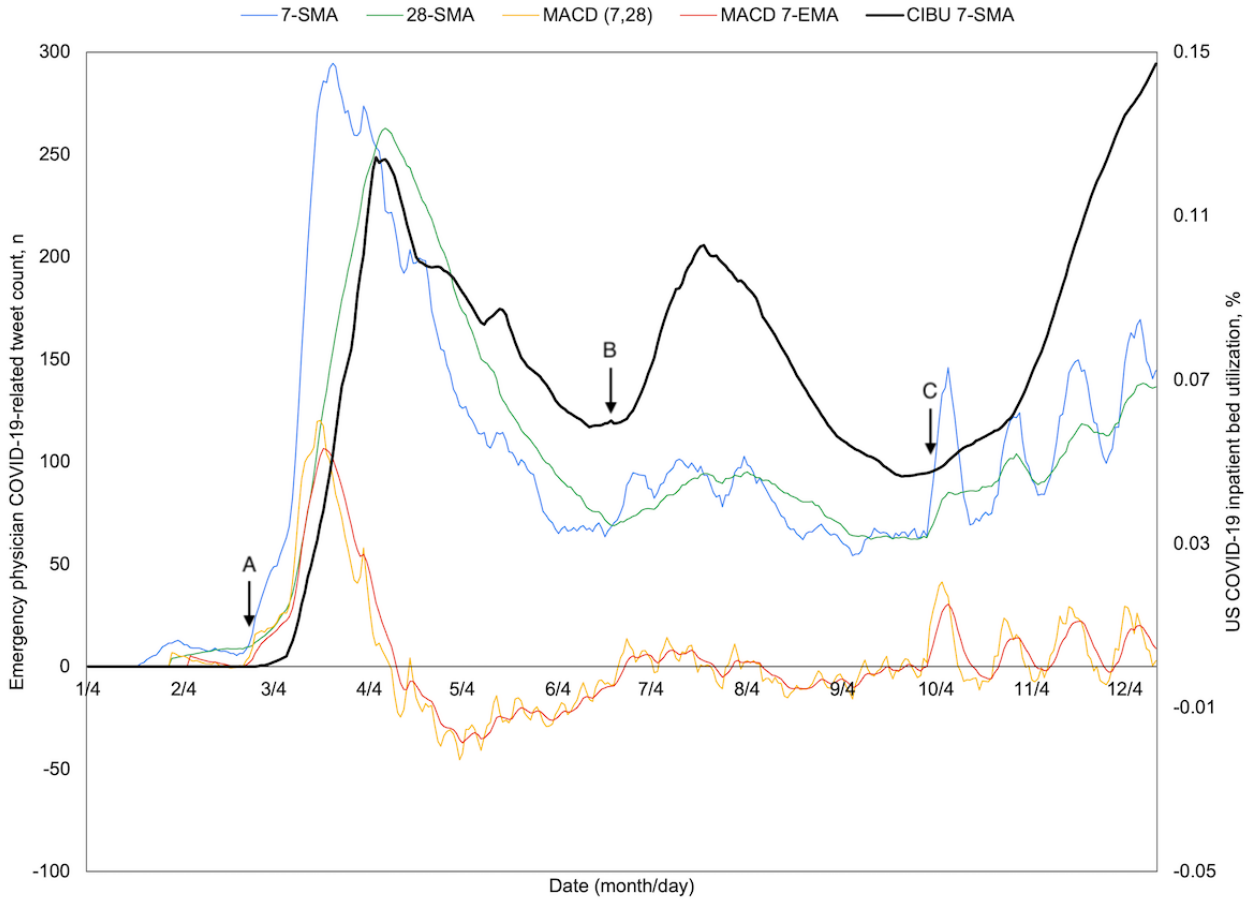


Over the full study period, three peaks emerged in both COVID-19–related discussion and CIBU, with the rise in tweets appearing to precede the corresponding rise in CIBU. This may be better appreciated with attention directed to where the 7-day moving average crosses above the 28-day moving average as a signal of topic momentum (Figure 4). After the first recorded domestic COVID-19 case on January 22, 2020, February 25 was the first such cross and preceded a period of sustained increase in CIBU from February 28 to an April 9 peak. The next occurrence was on June 22, occurring alongside the second

XSL•FO
RenderX

period of sustained increase in CIBU from June 15 to a peak on July 21. A brief cross on July 31 was short-lived, but the subsequent cross on September 13 was maintained and corresponded to a rise in COVID-19 hospital burden that started September 24 and continued through the remainder of the study period, with several episodic crosses seen thereafter. When the MACD is also considered, February 22 (Figure 4, point A) marks a cross above both the zero centerline as well as its 7-day moving average, 46 days before the April 9 peak in CIBU. The next such cross occurred on June 24 (Figure 4, point B), 27 days before the second peak, while the third surge in CIBU appeared to coincide with a first crossing on September 30 (Figure 4, point C).

Figure 4. Time series plot of percent US COVID-19 inpatient bed utilization (CIBU; right axis) and its 7-day simple moving average (CIBU 7-SMA; right axis) against the 7-SMA and 28-day simple moving average (28-SMA) of COVID-19–related emergency physician tweets (left axis). Also plotted are the tweet exponential moving average convergence/divergence oscillator (MACD; left axis) and its own 7-day exponential moving average signal line (MACD 7-EMA; left axis). Labels A through C demonstrate sustained crossover points for tweet volume, where both the 7-SMA overcomes the 28-SMA and the MACD 7-EMA turns positive and overcomes the MACD as indicators of momentum.



Because the breadth and diversity of the United States may obfuscate local trends, the four most contributory states of California, New York, Pennsylvania, and Texas were similarly plotted (Figures 5-8). All four experienced a spring signal and subsequent surge, although New York has been recognized among them as an early epicenter [44]. Only Texas appears to have had a sustained cross of the 7-day moving average above the 28-day moving average from June 18 to July 15. This notably preceded the only significant summer peak among these states, reaching a maximum CIBU of 20.5% on July 20; in comparison, California reached a second peak of 14.3% on July 25 while neither New York nor Pennsylvania exceeded 10% again before November. The mean time from the preceding cross of moving averages in COVID-19–related emergency physician tweets to peak CIBU across the four states and the nation was 45.0 (SD 12.7) days.

XSL•FO
RenderX

Figure 5. California time series plots of the 7-day simple moving average (7-SMA) in percent COVID-19 inpatient bed utilization (CIBU 7-SMA; right axis) against the 7-SMA and the 28-day simple moving average (28-SMA) of COVID-19–related emergency physician tweet count (left axis). Also plotted are the tweet exponential moving average convergence/divergence oscillator (MACD; left axis) and its own 7-day exponential moving average signal line (MACD 7-EMA; left axis).

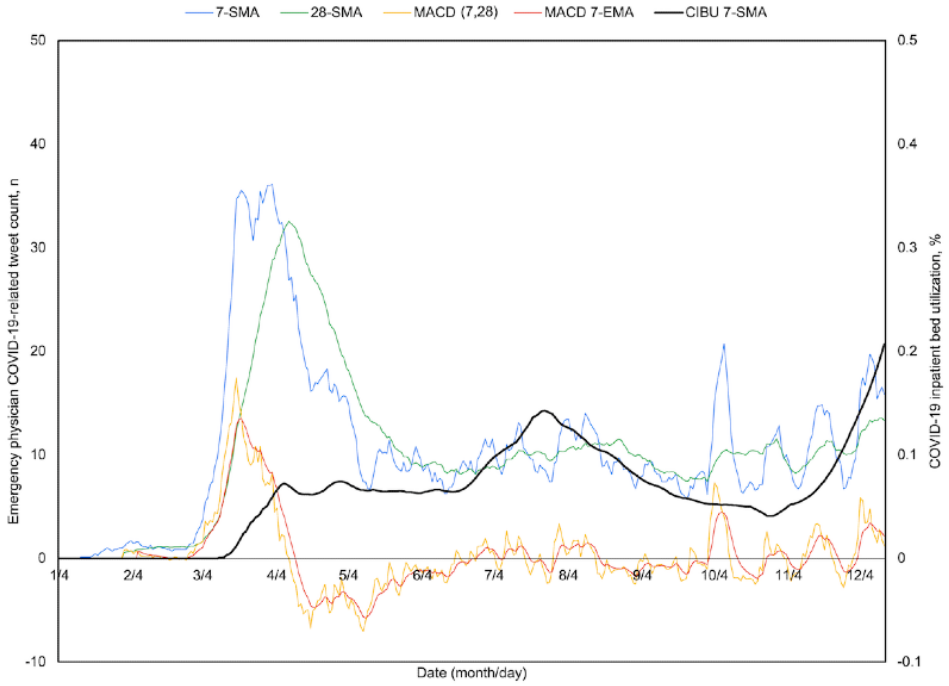
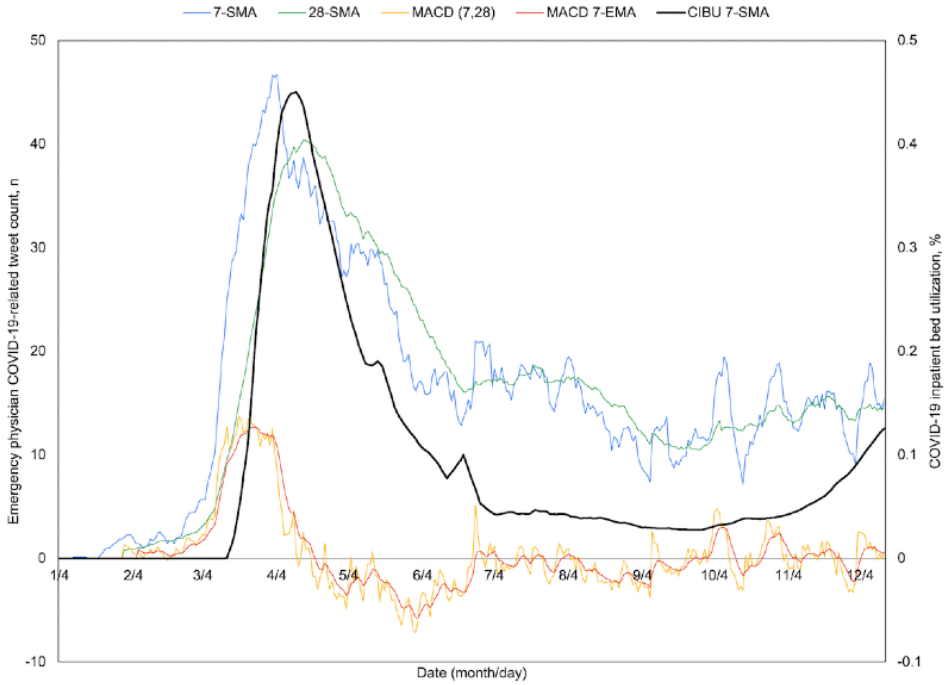


Figure 6. New York time series plots of the 7-day simple moving average (7-SMA) in percent COVID-19 inpatient bed utilization (CIBU 7-SMA; right axis) against the 7-SMA and the 28-day simple moving average (28-SMA) of COVID-19–related emergency physician tweet count (left axis). Also plotted are the tweet exponential moving average convergence/divergence oscillator (MACD; left axis) and its own 7-day exponential moving average signal line (MACD 7-EMA; left axis).



XSL•FO
RenderX

Figure 7. Pennsylvania time series plots of the 7-day simple moving average (7-SMA) in percent COVID-19 inpatient bed utilization (CIBU 7-SMA; right axis) against the 7-SMA and the 28-day simple moving average (28-SMA) of COVID-19–related emergency physician tweet count (left axis). Also plotted are the tweet exponential moving average convergence/divergence oscillator (MACD; left axis) and its own 7-day exponential moving average signal line (MACD 7-EMA; left axis).

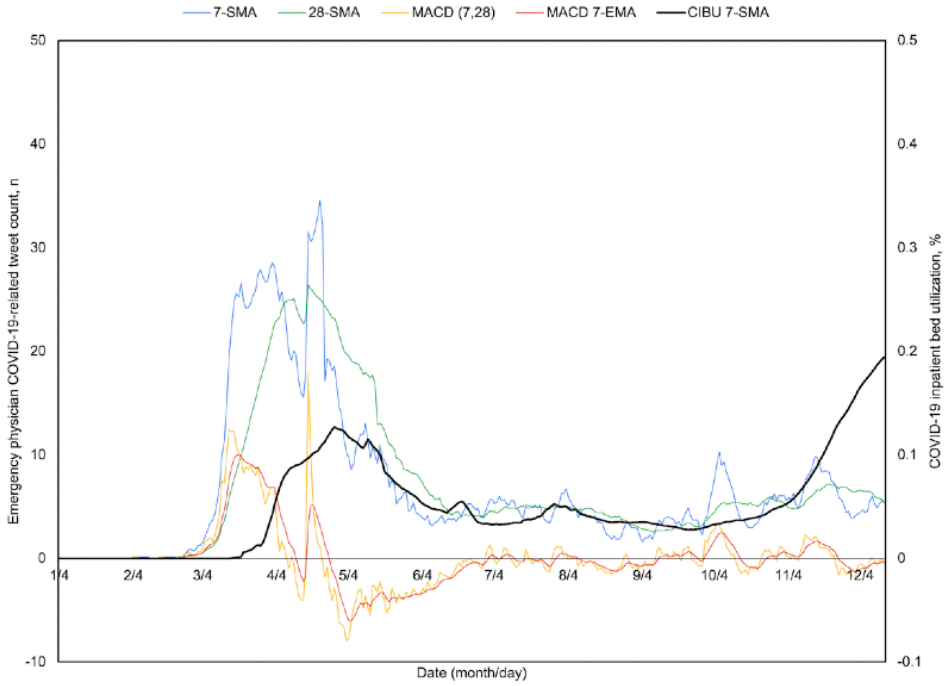
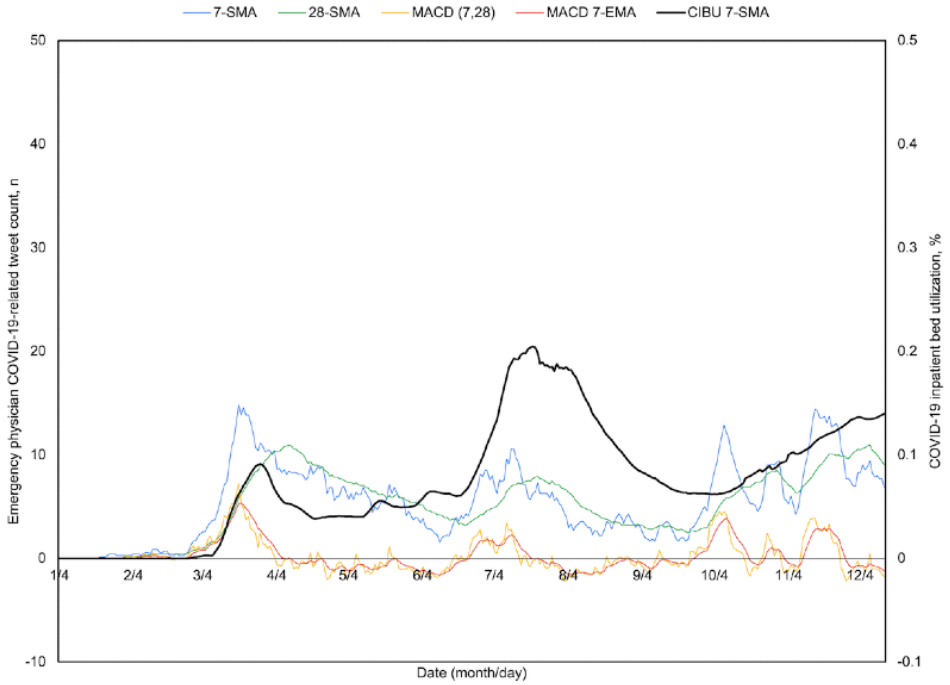


Figure 8. Texas time series plots of the 7-day simple moving average (7-SMA) in percent COVID-19 inpatient bed utilization (CIBU 7-SMA; right axis) against the 7-SMA and the 28-day simple moving average (28-SMA) of COVID-19–related emergency physician tweet count (left axis). Also plotted are the tweet exponential moving average convergence/divergence oscillator (MACD; left axis) and its own 7-day exponential moving average signal line (MACD 7-EMA; left axis).



Discussion

Principal Findings

Emergency physician engagement on Twitter has grown considerably since the start of the COVID-19 pandemic in both the topics raised and the sentiments conveyed. Furthermore, the analysis described here demonstrates conversations with increasing focus on pandemic response, clinical care, and epidemiology. That these topics correlate better with CIBU than simple case counts supports the idea that they may well serve as a kind of barometer of health care system strain. Momentum

XSL•FO
RenderX

in these conversations, in fact, as shown by crossings of tweet count moving averages, were shown to occur before key rises in CIBU, which may with future research lend itself to the larger effort of predicting surge based on multiple data streams.

The COVID-19 pandemic has required emergency physicians to adapt continually to an extraordinary inadequacy of resources. While strains on the ventilator supply received much attention [45], early limitations in testing and bed availability also created significant clinical challenges [46], let alone the mental health effects that are likely to be far-reaching [47]. In the case of personal protective equipment (PPE) shortages, many frontline providers resorted to individual means to acquire makeshift supplies, and some even turned to social media, as in the case of the #GetMePPE Twitter hashtag, in order to spur necessary action [48-50]. Despite that potential good, such public distress and debate from the medical frontline has, at times, spurred controversy and even real-world, professional repercussions [51].

With overwhelming caseloads and without clinical consensus, frontline physicians have been forced to decide between various treatment modalities based on unclear and, at times, contradictory information, with significant moral distress [52]. The effort to maintain appropriate patient care despite these unknowns, when faced with a need for resource rationing [53], is a *de facto* implementation of crisis standards of care. While contingency planning is situational and incorporates some aspects of triage practiced routinely in overcrowded emergency departments across the country, the formal triggering of crisis standards, and implicit divergence from conventional standards, enacts systematic change in protocols and care plans during a sustained period of large-scale strain [54-56].

Taken together, this retrospective look at emergency physician Twitter use suggests a new way of considering the pandemic surge, as emergency physician utilization of Twitter reached unprecedented highs. There are likely several reasons for this. The online community has been shown to provide psychological benefit, potentially exacerbated by the isolation faced in providing crisis care, and by a perceived collapse in trust in the existing infrastructure and policy guidance [57,58]. That *collaboration* was the only topic among 20 to have a compound sentiment range that did not cross zero may relate to this yearning for support. Still, positive mean sentiment scores among the vast majority of topics were unexpected, given recent work pertaining to general public perceptions of the pandemic [59]. There may well be some sway to a self-perceived personal and professional connection to the dominant issue of the day. While the root cause is undoubtedly complex and multifaceted, the increase in emergency physician engagement with social media is likely here to stay.

Whether emergency physicians online can truly act as an early indicator for policy makers remains to be seen, but the community is undoubtedly a subset of the broader pandemic response and is worth looking at more closely. Moving averages have been used to indicate movement in financial markets but are not true predictors of future trends. Given the potential for false signaling and the challenge in determining what constitutes a sustained or meaningful cross prospectively, derived crosses of the kind shown here will likely need corroboration from a variety of other metrics as well as comparison to other samples and controls. Even so, the idea that an indicator of a physician behavioral trend online may also signal momentum in real-world hospital bed utilization has clear implications for the future. This is particularly relevant when it is considered that the study group itself was sourced directly from followers of major professional organizations for whom early recognition of surge would empower a more coordinated and efficacious policy response.

To make such a tool operationally relevant, collaboration between government and private sector partners will likely be necessary to build adequate data pathways allowing for public health surveillance in real time. While emergent topic generation from a retrospective corpus of tweets is not a feasible option for rapid and predictive modeling, this work suggests that even a simple collection of tweets containing disease-specific text strings can nonetheless yield important, potentially meaningful information to inform resource allocation and other policy decisions. Ultimately, all disasters are media events, affecting both how and what information is conveyed. There is, therefore, no great leap of faith in acknowledging that social media, too, may have an important part to play. Future research must delve further into how such tools can one day be used in the early recognition of, and response to, health care strain of such magnitude.

Limitations

In holding to strict inclusion criteria, this study overlooked Twitter users who were not explicit in their self-identification as practicing emergency physicians. Emergency physicians were made the narrow focus of this work based on their key roles as clinical decision makers overseeing department throughput, but inclusion of nurses, technicians, and nonphysician midlevel providers may add breadth. Additionally, follower referral from within the sample may have introduced bias that could have been avoided by subsampling with some individuals selected for study participation and others only for referral [25]. Even so, a comprehensive list of emergency physicians on Twitter compiled in 2016 concluded that there were only 2234 such users around the globe [60]. Social media use has undoubtedly risen since, particularly with the influx of a growing number of emergency medicine residents [61], but the sample provided here does appear to be appropriately sized for this purpose.

Even so, this sample size was insufficient in both participant number and geographic spread to allow for more granular geographic analysis by city or county, although public health surveillance often occurs at this level [62,63]. Both demographic characteristics and spaciotemporal effects at the level of the individual participant have previously been shown to bias tweet sentiment and content, but these were not controlled for in this study [64,65].

Reliance on Twitter may itself limit generalizability, given its comparatively higher representation of young, urban, and minority users when compared with the general US population [66]. Only one social network platform was analyzed, and, insofar as it serves as a public forum, what medical professionals

XSL•FO
RenderX

say online does not necessarily correlate with what they think or feel [67]. The study group, however, was not aware of its participation in this research, thereby avoiding that influence on behavior [68]. Excluding reposted tweets may have overlooked certain sentiments and reactions. Additionally, LDA topic modeling depends not only on the size of the overall corpus but on the length of the individual documents themselves. Although methods such as aggregation into larger documents have been proposed in order to overcome tweet brevity [69], doing so would not have allowed for the temporal and user-specific analysis intended. Finally, care must be taken in interpreting relationships between variables, such as physician tweets and CIBU, when both variables have undergone averaging or smoothing, which can sometimes suggest correlation where none exists.

Conclusions

This work reveals both the opportunity and the pressing need to explore social media use by the emergency physician community as a means of anticipating surge needs. By acting as gatekeepers to the hospital, emergency physicians are uniquely positioned to act as early indicators of hospital surge, and finding methods such as Twitter usage, which can track and analyze these indicators, could be vital to future pandemic planning and response.

Conflicts of Interest

None declared.

References

1. World Health Organization (WHO). Twitter. 2020 Jan 04. URL: https://twitter.com/WHO/status/1213523866703814656 [accessed 2020-08-04]
2. Twitter Q3 2020 shareholder letter. US Securities and Exchange Commission. San Francisco, CA: Twitter, Inc; 2020 Oct 29. URL: https://www.sec.gov/Archives/edgar/data/1418091/000141809120000199/twtrq320ex991.htm [accessed 2020-12-10]
3. Hinton DM. Pfizer-BioNTech COVID-19 vaccine EUA letter of authorization. US Food and Drug Administration. Silver Spring, MD: US Food and Drug Administration; 2021 May 10. URL: https://www.fda.gov/media/144412/download [accessed 2020-12-11]
4. US Food and Drug Administration (US FDA). Twitter. 2020 Dec 11. URL: https://twitter.com/US_FDA/status/1337588046674489346 [accessed 2020-12-11]
5. Liu M, Caputi TL, Dredze M, Kesselheim AS, Ayers JW. Internet searches for unproven COVID-19 therapies in the United States. JAMA Intern Med 2020 Aug 01;180(8):1116-1118 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1764] [Medline: 32347895]
6. Chew C, Eysenbach G. Pandemics in the age of Twitter: Content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
7. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
8. Ioannidis JP, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. Int J Forecast 2020 Aug 25:1-16 [FREE Full text] [doi: 10.1016/j.ijforecast.2020.08.004] [Medline: 32863495]
9. Yang J. California and New York were both hit by Covid-19 early, but the results are very different. CNN. 2020 Apr 16. URL: https://www.cnn.com/2020/04/14/opinions/california-new-york-covid-19-coronavirus-yang/index.html [accessed 2021-01-17]
10. Secon H. England's coronavirus death rate is nearly 3 times higher than Ireland's. Proactive social-distancing measures made the difference. Business Insider. 2020 Apr 18. URL: https://www.businessinsider.com/england-ireland-coronavirus-response-compared-2020-4 [accessed 2020-01-17]
11. Wells P. California surpasses New York as centre of US Covid crisis. Financial Times. 2021 Jan 16. URL: https://www.ft.com/content/8013b83a-84ab-4a8e-b39a-2f28c4d710f4 [accessed 2021-01-17]
12. Carswell S. How did Ireland jump from low Covid base to world's highest infection rate? The Irish Times. 2021 Jan 16. URL: https://www.irishtimes.com/news/health/how-did-ireland-jump-from-low-covid-base-to-world-s-highest-infection-rate-1.4459429 [accessed 2021-01-17]
13. Stanley K, Lora M, Merjavy S, Chang J, Arora S, Menchine M, et al. HIV prevention and treatment: The evolving role of the emergency department. Ann Emerg Med 2017 Oct;70(4):562-572.e3. [doi: 10.1016/j.annemergmed.2017.01.018] [Medline: 28347557]
14. Cochrane DG. Perspective of an emergency physician group as a data provider for syndromic surveillance. MMWR Suppl 2004 Sep 24;53:209-214 [FREE Full text] [Medline: 15717394]
15. Greene J. Social media and physician learning: Is it all Twitter? Ann Emerg Med 2013 Nov;62(5):11A-13A. [doi: 10.1016/j.annemergmed.2013.09.013] [Medline: 24319781]
16. Roesslein J. Tweepy: Twitter for Python!. GitHub. 2020. URL: https://github.com/tweepy/tweepy [accessed 2020-12-11]

XSL•FO
RenderX

17. Joint statement on mask excuses in the ED. American College of Emergency Physicians. 2020 Jul 28. URL: https://www.acep.org/corona/COVID-19-alert/covid-19-articles/joint-statement-on-mask-excuses-in-the-ed/ [accessed 2020-12-11]
18. EM organizations. West Virginia University Libraries. 2020 Nov 05. URL: https://libguides.wvu.edu/EmergencyMedicine/ProfOrgs [accessed 2020-12-12]
19. Moore J. Professional Organizations-Emergency Medicine. Minneapolis, MN: University of Minnesota URL: https://tinyurl.com/pujrymmt [accessed 2020-12-12]
20. Emergency medicine specialty description. American Medical Association. URL: https://www.ama-assn.org/specialty/emergency-medicine-specialty-description [accessed 2020-12-12]
21. Counselman FL, Beeson MS, Marco CA, Adsit SK, Harvey AL, Keehbauch JN, 2016 EM Model Review Task Force. Evolution of the Model of the Clinical Practice of Emergency Medicine: 1979 to present. Acad Emerg Med 2017 Feb;24(2):257-264 [FREE Full text] [doi: 10.1111/acem.13137] [Medline: 27859987]
22. Shillcutt SK, Silver JK. Social media and advancement of women physicians. N Engl J Med 2018 Jun 14;378(24):2342-2345. [doi: 10.1056/nejmms1801980]
23. Safdar B, Ona Ayala KE, Ali SS, Seifer BJ, Hong M, Greenberg MR, et al. Inclusion of sex and gender in emergency medicine research-A 2018 update. Acad Emerg Med 2019 Mar;26(3):293-302 [FREE Full text] [doi: 10.1111/acem.13688] [Medline: 30637897]
24. Zhu JM, Pelullo AP, Hassan S, Siderowf L, Merchant RM, Werner RM. Gender differences in Twitter use and influence among health policy and health services researchers. JAMA Intern Med 2019 Dec 01;179(12):1726-1729 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4027] [Medline: 31609393]
25. Avrachenkov K, Neglia G, Tuholukova A. Subsampling for chain-referral methods. In: Proceedings of the International Conference on Analytical and Stochastic Modeling Techniques and Applications.: Springer International Publishing; 2016 Presented at: International Conference on Analytical and Stochastic Modeling Techniques and Applications; August 24-26, 2016; Cardiff, UK p. 17-31. [doi: 10.1007/978-3-319-43904-4_2]
26. Otterman S. 'I trust science,' says nurse who is first to get vaccine in US. The New York Times. 2020 Dec 14. URL: https://www.nytimes.com/2020/12/14/nyregion/us-covid-vaccine-first-sandra-lindsay.html [accessed 2020-12-18]
27. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: Review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]
28. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI URL: https://www.scinapse.io/papers/2099813784
29. Raghupathi V, Ren J, Raghupathi W. Studying public perception about vaccination: A sentiment analysis of tweets. Int J Environ Res Public Health 2020 May 15;17(10):3464 [FREE Full text] [doi: 10.3390/ijerph17103464] [Medline: 32429223]
30. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
31. How to use hashtags. Twitter. URL: https://help.twitter.com/en/using-twitter/how-to-use-hashtags [accessed 2021-01-14]
32. Parker A. Twitter's secret handshake. The New York Times. 2011 Jun 10. URL: https://www.nytimes.com/2011/06/12/fashion/hashtags-a-new-way-for-tweets-cultural-studies.html [accessed 2021-01-14]
33. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
34. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
35. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. In: Proceedings of the 8th ACM International Conference on Web Search and Data Mining.: Association for Computing Machinery; 2015 Presented at: 8th ACM International Conference on Web Search and Data Mining; February 2-6, 2015; Shanghai, China p. 399-408. [doi: 10.1145/2684822.2685324]
36. CDC Case Task Force. United States COVID-19 cases and deaths by state over time. Centers for Disease Control and Prevention. 2020 Dec 22. URL: https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36 [accessed 2020-12-23]
37. COVID-19 reported patient impact and hospital capacity by state timeseries. HealthData.gov. 2021. URL: https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh [accessed 2021-01-18]
38. Evaluation estimates. United States Census Bureau. URL: https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates.html [accessed 2020-12-22]
39. The World Factbook: Puerto Rico. Central Intelligence Agency. URL: https://www.cia.gov/the-world-factbook/countries/puerto-rico/ [accessed 2020-12-23]
40. Chong TT, Ng W. Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. Appl Econ Lett 2008 Nov 21;15(14):1111-1114. [doi: 10.1080/13504850600993598]

XSL•FO
RenderX

41. Paroli M, Sirinian MI. Predicting SARS-CoV-2 infection trend using technical analysis indicators. Disaster Med Public Health Prep 2021 Feb;15(1):e10-e14 [FREE Full text] [doi: 10.1017/dmp.2020.254] [Medline: 32674742]
42. Mishori R, Singh LO, Levy B, Newport C. Mapping physician Twitter networks: Describing how they work as a first step in understanding connectivity, information flow, and message diffusion. J Med Internet Res 2014 Apr 14;16(4):e107 [FREE Full text] [doi: 10.2196/jmir.3006] [Medline: 24733146]
43. Sprunt B. "Don't be afraid of it": Trump dismisses virus threat as he returns to White House. NPR. 2020 Oct 05. URL: https://www.npr.org/sections/latest-updates-trump-covid-19-results/2020/10/05/920412187/trump-says-he-will-leave-walter-reed-medical-center-monday-night [accessed 2021-02-20]
44. Thompson C, Baumgartner J, Pichardo C, Toro B, Li L, Arciuolo R, et al. COVID-19 outbreak - New York City, February 29-June 1, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1725-1729 [FREE Full text] [doi: 10.15585/mmwr.mm6946a2] [Medline: 33211680]
45. Ranney ML, Griffeth V, Jha AK. Critical supply shortages — The need for ventilators and personal protective equipment during the Covid-19 pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/nejmp2006141]
46. Margus C, Sondheim SE, Peck NM, Storch B, Ngai KM, Ho H, et al. Discharge in pandemic: Suspected Covid-19 patients returning to the emergency department within 72 hours for admission. Am J Emerg Med 2020 Aug 18:1-7 [FREE Full text] [doi: 10.1016/j.ajem.2020.08.034] [Medline: 33046303]
47. Muller AE, Hafstad EV, Himmels JPW, Smedslund G, Flottorp S, Stensland S, et al. The mental health impact of the COVID-19 pandemic on healthcare workers, and interventions to help them: A rapid systematic review. Psychiatry Res 2020 Nov;293:113441 [FREE Full text] [doi: 10.1016/j.psychres.2020.113441] [Medline: 32898840]
48. Artenstein AW. In pursuit of PPE. N Engl J Med 2020 Apr 30;382(18):e46. [doi: 10.1056/nejmc2010025]
49. He S, Ojo A, Beckman A, Gondi S, Gondi S, Betz M, et al. The story of #GetMePPE and GetUsPPE.org to mobilize health care response to COVID-19: Rapidly deploying digital tools for better health care. J Med Internet Res 2020 Jul 20;22(7):e20469 [FREE Full text] [doi: 10.2196/20469] [Medline: 32530813]
50. Lee E, Loh W, Ang I, Tan Y. Plastic bags as personal protective equipment during the COVID-19 pandemic: Between the devil and the deep blue sea. J Emerg Med 2020 May;58(5):821-823 [FREE Full text] [doi: 10.1016/j.jemermed.2020.04.016] [Medline: 32359911]
51. Doubek J. ER doctor says he walks into a "war zone" every day. NPR. 2020 Dec 17. URL: https://www.npr.org/sections/coronavirus-live-updates/2020/12/17/946910354/er-doctor-says-he-walks-into-a-war-zone-every-day [accessed 2020-12-30]
52. Hertelendy A, Ciottone G, Mitchell C, Gutberg J, Burkle F. Crisis standards of care in a pandemic: Navigating the ethical, clinical, psychological and policy-making maelstrom. Int J Qual Health Care 2021 Mar 03;33(1):mzaa094 [FREE Full text] [doi: 10.1093/intqhc/mzaa094] [Medline: 33128564]
53. Wynia MK. Crisis triage-Attention to values in addition to efficiency. JAMA Netw Open 2020 Dec 01;3(12):e2029326 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.29326] [Medline: 33315107]
54. Leider JP, DeBruin D, Reynolds N, Koch A, Seaberg J. Ethical guidance for disaster response, specifically around crisis standards of care: A systematic review. Am J Public Health 2017 Sep;107(9):e1-e9. [doi: 10.2105/ajph.2017.303882]
55. Hick JL, Barbera JA, Kelen GD. Refining surge capacity: Conventional, contingency, and crisis capacity. Disaster Med Public Health Prep 2013 Apr 08;3(S1):S59-S67. [doi: 10.1097/dmp.0b013e31819f1ae2]
56. Margus C, Sarin RR, Molloy M, Ciottone GR. Crisis standards of care implementation at the state level in the United States. Prehosp Disaster Med 2020 Sep 10;35(6):599-603. [doi: 10.1017/s1049023x20001089]
57. Mira JJ, Carrillo I, Guilabert M, Mula A, Martin-Delgado J, Pérez-Jover MV, SARS-CoV-2 Second Victim Study Group. Acute stress of the healthcare workforce during the COVID-19 pandemic evolution: A cross-sectional study in Spain. BMJ Open 2020 Nov 06;10(11):e042555 [FREE Full text] [doi: 10.1136/bmjopen-2020-042555] [Medline: 33158839]
58. Bachem R, Tsur N, Levin Y, Abu-Raiya H, Maercker A. Negative affect, fatalism, and perceived institutional betrayal in times of the coronavirus pandemic: A cross-cultural investigation of control beliefs. Front Psychiatry 2020;11:589914 [FREE Full text] [doi: 10.3389/fpsyt.2020.589914] [Medline: 33192739]
59. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: Aentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
60. Riddell J, Brown A, Kovic I, Jauregui J. Who are the most influential emergency physicians on Twitter? West J Emerg Med 2017 Feb;18(2):281-287 [FREE Full text] [doi: 10.5811/westjem.2016.11.31299] [Medline: 28210365]
61. Haas MRC, Hopson LR, Zink BJ. Too big too fast? Potential implications of the rapid increase in emergency medicine residency positions. AEM Educ Train 2020 Feb;4(Suppl 1):S13-S21 [FREE Full text] [doi: 10.1002/aet2.10400] [Medline: 32072104]
62. Enanoria WT, Crawley AW, Hunter JC, Balido J, Aragon TJ. The epidemiology and surveillance workforce among local health departments in California: Mutual aid and surge capacity for routine and emergency infectious disease situations. Public Health Rep 2014 Nov 01;129(6_suppl4):114-122. [doi: 10.1177/00333549141296s415]
63. Lynch CJ, Gore R. Short-range forecasting of COVID-19 during early onset at county, health district, and state geographic levels using seven methods: Comparative forecasting study. J Med Internet Res 2021 Mar 23;23(3):e24925 [FREE Full text] [doi: 10.2196/24925] [Medline: 33621186]

XSL•FO
RenderX

64. Gore RJ, Diallo S, Padilla J. You are what you tweet: Connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
65. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
66. Smith A, Brenner J. Twitter use 2012. Pew Research Center. Washington, DC: Pew Research Center; 2012 May 31. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2021-04-09]
67. DeCamp M, Koenig TW, Chisolm MS. Social media and physicians' online identity crisis. JAMA 2013 Aug 14;310(6):581-582 [FREE Full text] [doi: 10.1001/jama.2013.8238] [Medline: 23942675]
68. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
69. Hong L, Davison B. Empirical study of topic modeling in Twitter. In: Proceedings of the 1st Workshop on Social Media Analytics.: Association for Computing Machinery; 2010 Presented at: 1st Workshop on Social Media Analytics; July 25, 2010; Washington, DC p. 80-88. [doi: 10.1145/1964858.1964870]

Abbreviations

AAEM: American Academy of Emergency Medicine
ACEP: American College of Emergency Physicians
API: application programming interface
CIBU: COVID-19 inpatient bed utilization
FDA: Food and Drug Administration
LDA: latent Dirichlet allocation
MACD: moving average convergence/divergence oscillator
PPE: personal protective equipment
SAEM: Society for Academic Emergency Medicine
VADER: Valence Aware Diction and Sentiment Reasoner
WHO: World Health Organization

Edited by C Basch; submitted 08.03.21; peer-reviewed by S Vilendrer, K Goniewicz, P Chai, R Gore; comments to author 25.03.21; revised version received 14.04.21; accepted 23.04.21; published 14.07.21.

Please cite as:
Margus C, Brown N, Hertelendy AJ, Safferman MR, Hart A, Ciottone GR
Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study
J Med Internet Res 2021;23(7):e28615
URL: https://www.jmir.org/2021/7/e28615
doi:10.2196/28615
PMID:34081612

©Colton Margus, Natasha Brown, Attila J Hertelendy, Michelle R Safferman, Alexander Hart, Gregory R Ciottone. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 14.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

COVID-19 Information Dissemination Using the WeChat Communication Index: Retrospective Analysis Study

Zina Fan[1,2], MA; Wenqiang Yin[1,3*], PhD; Han Zhang[1,2], MA; Dandan Wang[1,2], MA; Chengxin Fan[1,2], MA; Zhongming Chen[1,3], PhD; Jinwei Hu[1,3], MA; Dongping Ma[1,3], PhD; Hongwei Guo[1,2*], MA

[1]School of Management, Weifang Medical University, Weifang, Shandong, China

[2]"Health Shandong" Collaborative Innovation Center for Severe Social Risk Prediction and Governance, Weifang, Shandong, China

[3]Collaborative Innovation Center of Social Risks Governance in Health, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Hongwei Guo, MA
School of Management
Weifang Medical University
No. 7166 Baotong West Street
Weifang, Shandong, 261053
China
Phone: 86 05368462573
Email: guohongww@126.com

Abstract

Background: The COVID-19 outbreak has tremendously impacted the world. The number of confirmed cases has continued to increase, causing damage to society and the economy worldwide. The public pays close attention to information on the pandemic and learns about the disease through various media outlets. The dissemination of comprehensive and accurate COVID-19 information that the public needs helps to educate people so they can take preventive measures.

Objective: This study aimed to examine the dissemination of COVID-19 information by analyzing the information released by the official WeChat account of the *People's Daily* during the pandemic. The most-read COVID-19 information in China was summarized, and the factors that influence information dissemination were studied to understand the characteristics that affect its dissemination. Moreover, this was conducted in order to identify how to effectively disseminate COVID-19 information and to provide suggestions on how to manage public opinion and information governance during a pandemic.

Methods: This was a retrospective study based on a WeChat official account. We collected all COVID-19–related information, starting with the first report about COVID-19 from the *People's Daily* and ending with the last piece of information about lifting the first-level emergency response in 34 Chinese provinces. A descriptive analysis was then conducted on this information, as well as on Qingbo Big Data's dissemination index. Multiple linear regression was utilized to study the factors that affected information dissemination based on various characteristics and the dissemination index.

Results: From January 19 to May 2, 2020, the *People's Daily* released 1984 pieces of information; 1621 were related to COVID-19, which mainly included headline news items, items with emotional content, and issues related to the pandemic's development. By analyzing the dissemination index, seven information dissemination peaks were discerned. Among the three dimensions of COVID-19 information—media salience, content, and format—eight factors affected the spread of COVID-19 information.

Conclusions: Different types of pandemic-related information have varying dissemination power. To effectively disseminate information and prevent the spread of COVID-19, we should identify the factors that affect this dissemination. We should then disseminate the types of information the public is most concerned about, use information to educate people to improve their health literacy, and improve public opinion and information governance.

(J Med Internet Res 2021;23(7):e28563) doi:10.2196/28563

KEYWORDS

COVID-19; information dissemination; People's Daily; Chinese news; public health and communication; media salience; WeChat

XSL•FO
RenderX

Introduction

The COVID-19 outbreak has impacted the world tremendously, and the number of confirmed cases has only continued to increase, causing damage to society and the economy worldwide [1,2]. The public pays close attention to information about the epidemic and learns about the disease through various media outlets [3]. However, as the internet develops, the spread of pandemic-related information continues to become faster and more complex [4-6]. On the one hand, the dissemination of comprehensive and accurate COVID-19 information that the public needs helps to educate people so they can take preventive measures [7]. The Singapore government, for example, immediately established a national WhatsApp account during the outbreak's early stage to inform people living in Singapore about relevant government updates and initiatives. Over 635,000 people subscribed to the channel to receive these messages [8]. In another example, a Vietnamese dancer choreographed a dance about how to carefully wash one's hands and even started a dance challenge on TikTok. This dance challenge video went viral, resulting in millions of people learning about essential handwashing steps. The dissemination of this information played a critical role in fighting against the spread of COVID-19 [9]. On the other hand, the spread of rumors and false information is also accelerating [10,11]. If we cannot provide timely knowledge to effectively guide and manage the spread of public opinion and rumors on the internet, the consequences will be disastrous [12]. Therefore, it is necessary to research and analyze the dissemination of COVID-19 information in order to help control the disease and disseminate pandemic-related information that the public needs, while ensuring its correct and comprehensive transmission (ie, preventing deviations, omissions, and even errors in the transmission process) [13-15]. In this era of globalization, it may be difficult to prevent the spread of COVID-19, but the most effective way to prevent panic among people is to provide reliable information that meets the public's needs [16], information from a scientific viewpoint with a preventive attitude toward COVID-19 [17]. The first step in this effort is to study communication related to COVID-19.

WeChat is an active and important app in China [18]. WeChat official accounts disseminate information quickly and conveniently. They are, therefore, a significant means of disseminating health information [19,20]. To evaluate and measure the dissemination effect of WeChat official accounts, Qingbo Big Data Technology Co Ltd in Beijing, China, has researched and developed the WeChat Communication Index (WCI). Qingbo Big Data Technology Co Ltd is a company that provides research and development of new media big data evaluation systems, influence standards, and other big data services. It serves the Chinese government as well as China's top news media outlets (eg, Xinhua News Agency, *People's Daily*, and China National Radio), including internet companies [21]. This company has a respectable reputation, and the WCI that it developed has become one of the most rigorous standards for measuring the disseminating power of WeChat [22,23]. The WCI contains four main indicators concerning spread rates (ie, the spread rate of the whole article, the average spread rate of each article, the title spread rate, and the peak spread rate), eight secondary indicators, and a set of calculation formulas for the standardized scores. Specifically, the higher the WCI value, the better the dissemination effect. During the COVID-19 outbreak in China, the WCI of the *People's Daily* WeChat official account was ranked the highest among all the WeChat official accounts. This is also China's most representative official media [24]. Thus, it is an important channel for the Chinese people to obtain official COVID-19 information [25]. Qingbo Big Data not only recorded all of this information but also used the WCI formula to calculate and publish the dissemination index for each piece of information. By analyzing these pieces of information and their dissemination indices, we can determine significant details concerning the spread of COVID-19 information, especially the trend and influencing factors of COVID-19 information dissemination, which has important value for public health. This study contributes to the dissemination of COVID-19 information in a scientifically accurate manner and to helping people receive public health information more effectively, so they may be able to take preventive measures and control the spread of COVID-19. Moreover, in the era of globalization with the rapid development of the internet, this research can also provide a reference for relevant research in other countries using data from social media platforms, such as Twitter and Facebook. Additionally, this study can provide a reference for other countries to understand the trend of human behavior when consuming COVID-19 information, which is conducive to understanding public opinion during a pandemic, enhancing the ability of information governance during public health emergencies.

The aim of this study was to examine the dissemination of COVID-19 information. This included analyzing the status of the information released by the *People's Daily* official WeChat account. This took place during the pandemic, while researching the dissemination trend of China's COVID-19 information. Additionally, it involved analyzing the influencing factors and the key elements of COVID-19 information dissemination, while discussing information governance during a pandemic.

Methods

Data Collection

This study's data were derived from Qingbo Big Data's platform, which includes all COVID-19 information released by the *People's Daily* official WeChat account and the dissemination index of each piece of information. The research cycle covered a total of 105 days, exporting all COVID-19 information from January 19, 2020, which is when the *People's Daily* first released COVID-19 information, to May 2, 2020, which is when China's 34 provinces (ie, autonomous regions and municipalities) completely lifted the first-level response to public health emergencies [26]. This period was chosen because it covers the entire duration of the COVID-19 pandemic, from the mainstream media's first report to the subsequent development and normalization of the COVID-19 pandemic. The public experienced *information scarcity*, caused by few media reports being released at first, which was followed by an *information explosion* due to the media's collective focus. Finally, a public opinion vortex was caused by the surplus of

XSL•FO
RenderX

COVID-19 information [27]. This period comprehensively covers the whole process of COVID-19's information dissemination, including each stage of communication [28].

Inclusion and Exclusion Criteria

This study included information that was related to COVID-19 and published within the study's established period. Any information that did not meet these criteria was excluded. All the information about COVID-19 that was released by the official WeChat account of the *People's Daily* was included in this study. Specifically, out of the 1984 pieces of information that the account released, 1621 items were related to COVID-19.

Statistical Analysis

This study used SPSS, version 21.0 (IBM Corp), to conduct its statistical analyses. Descriptive analysis was used to describe the basic characteristics of the COVID-19 information that was released by the *People's Daily*, and it describes the information content with a higher dissemination power, according to the dissemination trend chart in the Results section. A multiple stepwise regression analysis was used to analyze the factors that affected COVID-19 information dissemination. Dummy variables were manually established for the multi-categorical disordering of variables, where the dummy variables were grouped under the same factor in the same block, then "ENTER" was selected as part of the inclusion method to ensure that these dummy variables entered and exited at the same time. Other continuous variables and binary variables were grouped into another block, and the inclusion method was "STEPWISE." Then, a multiple linear regression was performed after each setting. The model's goodness of fit was checked as well as whether there was a collinearity problem.

Ethics Approval

As per the protocol at Weifang Medical University, China, the Institutional Review Board does not review studies that do not involve human subjects. As this study did not include human subjects, ethics approval was not required.

Data Sharing

The data sets used and analyzed during this study are available from the corresponding author on reasonable request.

Results

COVID-19 Information Released by the People's Daily

From January 19 to May 2, 2020, the *People's Daily* released 1621 pieces of information related to COVID-19 (Table 1). In terms of media salience, this included 1129 headline news items (69.65%) and 492 nonheadline items (30.35%), which refer to prominence. This can also be divided into 1023 emotional information items (63.11%) and 598 neutral information items (36.89%), which refer to valence. The number of words for each information item was primarily between zero and 500. In terms of content, the main theme concerned the development of COVID-19 (410/1621, 25.29%). Meanwhile, the sources were primarily government agencies (1478/1621, 91.18%), and most of the information was original content from the *People's Daily*, including 923 items (56.94%). In terms of the information's format, the release time was primarily between 6:01 AM and 6 PM, and it was mostly presented as *text + pictures* (957/1621, 59.04%). Furthermore, the number of fonts and colors ranged between 1 and 6.

XSL•FO
RenderX

Table 1. Characteristics of COVID-19 information released by the *People's Daily*.

Dimensions and variables	Categories	Information items (N=1621), n (%)	WeChat Communication Index, mean (SD)
Media salience			
Prominence			
	Headline news items	1129 (69.65)	90,457.35 (3744.00)
	Nonheadline items	492 (30.35)	86,747.28 (1694.76)
Valence			
	Emotional information items	1023 (63.11)	89,765.96 (3898.38)
	Neutral information items	598 (36.89)	87,534.51 (2204.58)
Attention (number of words per piece of information)			
	0-500	1005 (62.00)	89,467.22 (3700.27)
	501-1000	344 (21.22)	88,061.71 (3231.44)
	>1000	272 (16.78)	88,988.84 (3540.65)
Informational content			
Theme			
	Policy and planning	241 (14.87)	88,620.69 (3110.21)
	Treatment and research	230 (14.19)	90,383.53 (4233.89)
	Initiative and mobilization	148 (9.13)	90,527.20 (3977.95)
	Stories about fighting COVID-19	294 (18.14)	90,492.20 (3624.73)
	Current prevalence status	410 (25.29)	86,919.27 (1909.89)
	Epidemic prevention knowledge	270 (16.66)	89,448.48 (3675.55)
	Other themes	28 (1.73)	88,436.71 (2766.13)
Source			
	Government	1478 (91.18)	88,861.01 (3455.07)
	Civil organization	2 (0.12)	91,606.40 (6029.22)
	Enterprise	104 (6.42)	90,943.57 (4243.16)
	Personal	28 (1.73)	92,992.48 (4356.11)
	Medical institution	5 (0.31)	92,970.53 (5142.01)
	Research organization	2 (0.12)	89,373.20 (4850.61)
	Other sources	2 (0.12)	93,718.10 (8885.36)
Originality			
	Original	923 (56.94)	89,143.03 (3434.57)
	Nonoriginal	698 (43.06)	89,016.80 (3854.00)
Format			
Time of release			
	12:01 AM-6 AM	5 (0.31)	91,673.75 (4746.00)
	6:01 AM-12 PM	695 (42.87)	88,494.94 (3230.48)
	12:01 PM-6 PM	521 (32.14)	89,622.55 (3909.18)
	6:01 PM-12 AM	400 (24.68)	89,392.62 (3710.13)
Vividness			
	Text	241 (14.87)	86,929.55 (2236.30)
	Text + pictures	957 (59.04)	88,964.66 (3614.09)
	Text + video	164 (10.12)	90,508.84 (3821.82)

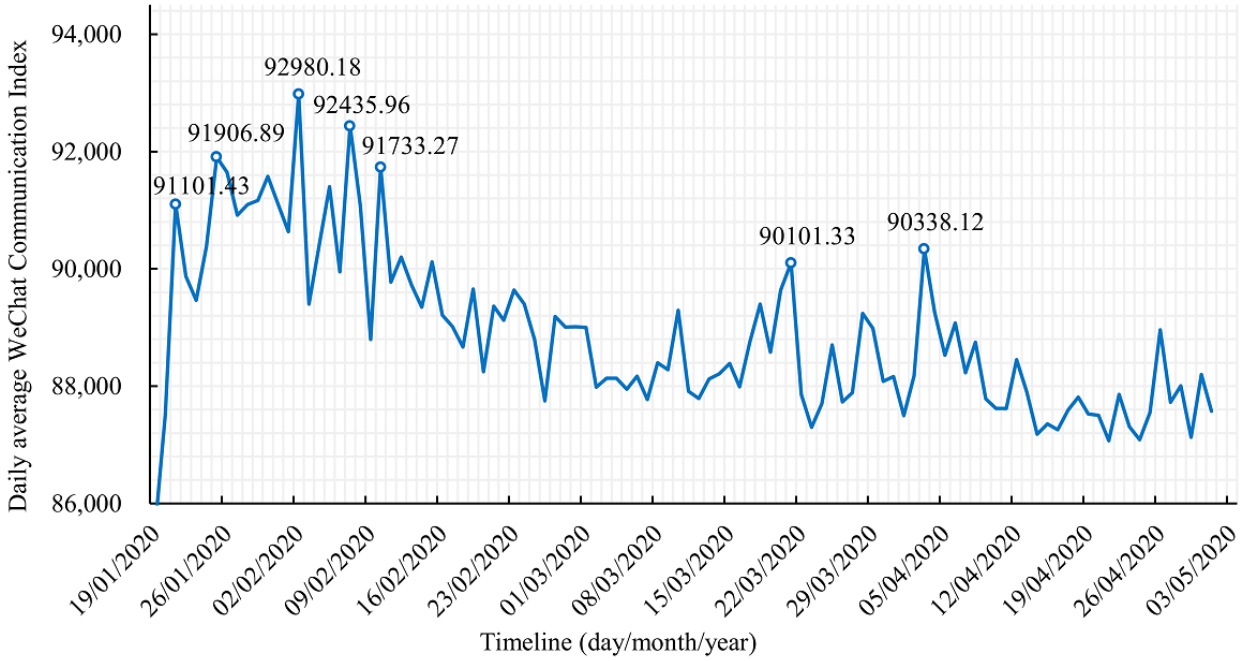
XSL•FO
RenderX

Dimensions and variables	Categories	Information items (N=1621), n (%)	WeChat Communication Index, mean (SD)
	Text + pictures + video	259 (15.98)	90,656.73 (3444.72)
Number of fonts			
	1 or 2	689 (42.50)	89,066.59 (3456.92)
	3 or 4	908 (56.01)	89,058.91 (3717.90)
	>4	24 (1.48)	90,849.01 (4173.84)
Number of colors			
	1 or 2	1360 (83.90)	88,903.74 (3579.23)
	3 or 4	252 (15.55)	90,090.93 (3721.28)
	>4	9 (0.56)	88,972.38 (2491.80)

Analysis of Informational Trends in the People's Daily

During the COVID-19 epidemic period, the *People's Daily* released several pieces of information related to COVID-19 every day, and the dissemination index for each piece of information represents its dissemination effect. Based on the time series analysis of the average dissemination index for all of the information released daily, we created a chart depicting the information dissemination trends. According to this chart, there were seven peaks of information dissemination from January 19, 2020, to 12 AM on May 2, 2020 (Figure 1). The analysis of these peaks and the content related to key events that occurred during these peaks revealed numerous insights.

Figure 1. Dissemination of COVID-19 information over time.



The first peak in the dissemination of COVID-19 information occurred around January 21 (mean dissemination index of 91,101.43). The *People's Daily* and other media began to report on COVID-19, which aroused public attention; in addition, the National Health Commission released the first announcement in 2020, which categorized COVID-19 as a Class B infectious disease, while adopting management measures as per the guidelines for more serious Class A infectious diseases. Specifically, when Zhong Nanshan confirmed that COVID-19 could be transmitted from person to person, the first peak of information dissemination was triggered. The second peak occurred around January 26, when the news reported that many new coronaviruses had been found through an epidemiological retrospective analysis [29]. The third peak occurred on February 2, when the news reported that the construction of Huoshenshan Hospital had been completed and the hospital was taken over by the Chinese People's Liberation Army. The dissemination index on that day was 92,980.18. The fourth peak of information dissemination occurred on February 7, when Li Wenliang passed away, arousing strong public concern. The *People's Daily* released five news items to commemorate him, and the State Supervision Commission appointed a team to investigate his death. The fifth peak happened on February 10, when the medical team's support for Wuhan reached its peak, and nearly 6000 medical workers arrived in Wuhan, the largest number of medical staff to arrive in Wuhan since the start of the COVID-19

XSL•FO
RenderX

epidemic. The sixth peak occurred on March 21, when no new cases had been reported in either Wuhan or the Hubei Province for the first time. In fact, there were no increases in cases for the first time in 31 provinces or in Xinjiang Production and Construction Corps. The news that the epidemic situation had improved aroused public concern. The seventh peak occurred around April 3, which was a period of national mourning. The *People's Daily* released some messages in memory of the dead, which triggered an emotional resonance among the public and led to a peak of information dissemination (mean dissemination index of 90,338.12).

Since then, the information dissemination index gradually decreased. Experts on infectious diseases and other public health emergencies found that the public have a strong curiosity and demand for information when they have a low awareness of diseases [30]. The distribution of peaks of information dissemination during this time reflects the public's psychological needs during an epidemic, illustrating that the public will seek the truth about an event at its early stage. At the middle stage, they will question some of the events that occurred, and at the epidemic's late stage, due to its normalization, the public's concern will gradually decline. The whole process presents a regular development trend [31].

Analysis of Factors Affecting the Dissemination of COVID-19 Information

The dimensions of media salience, information content, and format contained 10 factors. A multiple stepwise regression analysis was conducted with these 10 factors serving as independent variables and the dissemination indices as dependent variables. We found that eight variables entered the equation: prominence, valence, attention, theme, source, vividness, originality, and the number of fonts (Table 2). As the tolerance of the equation was greater than 0.1 and the variance inflation factor was less than 10, we could determine that no collinearity existed between independent variables (ie, the credibility of the multiple linear regression was high) [32]. As shown in Table 2, in terms of media salience, the dissemination of headline news items was better than that of nonheadline items, while the dissemination of emotional information was better than that of neutral information. Further, the fewer words that a piece had, the better its dissemination. In terms of informational content, the following variables affected the dissemination of COVID-19 information ($P<.05$): theme, source, and originality. In terms of the information's format, its vividness affected its dissemination; specifically, this pertains to whether information was presented by text, picture, or video. The number of fonts utilized also affected dissemination, as more font variety equated to a superior dissemination effect. However, the release time and number of colors had no significant impact on the dissemination of COVID-19 information.

XSL•FO
RenderX

Table 2. Analysis of factors affecting the dissemination of COVID-19 information.

Dimensions and variables	Regression coefficient (SE)	Standardized regression coefficient	*t* value	*P* value
Media salience				
Constant	64,378.147 (10,962.848)		5.872	<.001
Prominence	2150.089 (185.234)	0.273	11.607	<.001
Valence	2523.999 (199.787)	0.336	12.633	<.001
Attention	–0.151 (0.071)	–0.048	–2.122	.03
Informational content				
Theme (other theme = 0)				
Policy and planning	1121.800 (593.446)	0.110	1.890	.06
Treatment and research	2079.100 (592.525)	0.200	3.509	<.001
Initiative and mobilization	1777.394 (610.512)	0.141	2.911	.004
Stories of fighting COVID-19	1253.960 (584.977)	0.133	2.144	.03
Current prevalence status	421.651 (592.767)	0.051	0.711	.48
Epidemic prevention knowledge	1641.571 (592.536)	0.169	2.770	.006
Source (other sources = 0)				
Government	4943.180 (2090.008)	0.387	2.365	.02
Civil organization	3286.498 (2949.465)	0.032	1.114	.27
Enterprise	4174.864 (2104.164)	0.283	1.984	.047
Personal	2238.679 (2160.768)	0.081	1.036	.30
Medical institution	3091.545 (2467.104)	0.047	1.253	.21
Research organization	5957.447 (2946.397)	0.058	2.022	.04
Originality	481.217 (201.708)	0.066	2.386	.02
Format				
Vividness (text = 0)				
Text + pictures	772.974 (265.608)	0.105	2.910	.004
Text + video	788.653 (366.266)	0.066	2.153	.03
Text + pictures + video	758.213 (332.765)	0.077	2.279	.02
Number of fonts	318.600 (90.585)	0.096	3.517	<.001

Discussion

Principal Findings

Based on results of the WCI, this study examined the COVID-19 information that was first spread by the *People's Daily*. This was conducted by analyzing the information's media salience, content, and format. Thereafter, we summarized the COVID-19 information that was popular during the pandemic period. Seven information dissemination peaks were discerned. We further analyzed the influencing factors and key elements of COVID-19 information dissemination. We found that among the three dimensions of COVID-19 information—media salience, content, and format—eight factors affected the spread of COVID-19 information.

During the COVID-19 pandemic period, most of the information released by the *People's Daily* focused on the pandemic's development, with daily announcements about the number of confirmed cases, which kept the public abreast of related developments and helped dispel rumors. The public WeChat account of the *People's Daily* mostly took the form of headlines upon releasing COVID-19 information, and most items had positive emotional tendencies. It seems that extensive positive reports played a role in stabilizing people's emotions [33]. Most of the information came from the national and local health commissions, which ensured that the *People's Daily* portrayed authority and fairness [34]. The *People's Daily* is the most authoritative official media channel in China and, as such, the WCI of its official WeChat account ranked the highest during the epidemic. It was the responsibility of official media outlets to release relevant information about COVID-19 in a timely manner to spread relevant knowledge. This helped the public to have an awareness of the pandemic and improved the situation in China [35]. This shows the network media's power when it comes to disseminating COVID-19 information [36].

XSL•FO
RenderX

By analyzing the trend regarding the average daily dissemination index of information, as well as content and key events, we found that there were seven peaks of information transmission during the COVID-19 epidemic period, which began on the following dates in 2020: January 21, January 26, February 2, February 7, February 10, March 21, and April 3. These peaks were primarily concentrated on the following issues: the media's first reported coverage about COVID-19, epidemiological studies that confirmed the existence of human-to-human disease transmission, and government measures to manage the epidemic. These findings are consistent with research conducted by Liu Lanlan, who found that the dissemination peaks corresponded to January 26 and other dates similar to those found in this study, and information mainly concentrated on Zhong Nanshan's confirmation of COVID-19's transmission via person-to-person contact and Li Wenliang's commemoration, among other key events [31]. The WCI's peak value shows where the public's attention was focused and their psychological need for epidemic information. Thus, the mainstream media should release COVID-19 information in a targeted manner and disseminate the type of information that the public values the most [37]. It should also improve the timeliness of information release, educate people on preventive measures more effectively, and improve people's health literacy so they can better manage COVID-19 [38].

Our research shows that media salience, information content, and format affect COVID-19 information dissemination, specifically in terms of eight variables: prominence, valence, attention, theme, source, vividness, originality, and number of fonts. However, the time that the information was released and its number of colors had no effect on dissemination. These results suggest that different types of epidemic information disseminate differently; therefore, we must consider the factors affecting information dissemination and choose the best dissemination practices during the process of COVID-19 prevention and control to ensure that information is spread effectively and in a responsible manner. This would thereby prevent the spread of rumors and false information, improve the information's scientific and authoritative reputation [39], increase the ability to manage information and public opinion during epidemics, and contribute to a comprehensive public health system [40].

Strengths and Limitations

This study has several strengths regarding its design and analysis. First, the selected research objects were relatively representative. The WCI of the *People's Daily* official WeChat account showed that it was top-ranked among all WeChat official accounts, and this account is an important channel for the Chinese people to obtain official COVID-19 information. Second, based on big data technology, the WCI obtains data from all WeChat users (more than one billion) [41], resulting in a large sample size and high accuracy; therefore, the research findings are representative. Finally, the findings have practical significance; for example, this study found that headline news item dissemination has more impact than nonheadline items. These findings could inform the accurate dissemination of COVID-19 information and knowledge in the future. However, several limitations should be considered when interpreting the results. Because of the large number of potential factors and limited related research, we selected the three most commonly used dimensions—constituting 10 variables—from existing research, which may limit the research results. In addition, this study mainly analyzed the most popular social media platform in China. Although this study serves as an example for other countries to study COVID-19 information dissemination, it may not be completely applicable to other countries. This limits the inferences one may make from these results. Future studies can use more comprehensive potential factors to verify or enrich the conclusions of this study.

Conclusions

Currently, only a few indicators have been explored to quantify the effect of COVID-19 information dissemination. It is difficult to measure the spread of information when using quantitative data. It is even more complex as it occurs in conjunction with the unexpected and uncertain nature of COVID-19. This leads to only a few studies in the literature that have attempted to understand the effect and measurement of COVID-19 information dissemination. Thus, to fill this gap, we measured the dissemination of COVID-19–related information using WCI data, integrating the areas of public health and communication science (ie, media salience), to determine the factors affecting information dissemination. This study showed that the different types of pandemic-related information had varying dissemination power. It also revealed the factors that affected this dissemination: media salience, information content, and information format. These findings suggest that for effective dissemination of COVID-19 information, we should pay attention to the factors that affect this dissemination, which include the theme, source, and originality of COVID-19 information. Furthermore, we should disseminate the type of information that the public is most concerned about, choose those dissemination practices that effectively spread information in a responsible manner to curtail misinformation, regulate fear among the public, and improve information governance. Further research is needed to confirm the findings of this study and to validate the impact of these findings on the effect of COVID-19 information dissemination in other countries.

Acknowledgments

This study was supported by the Development Plan of Youth Innovation Team in Colleges and Universities of Shandong Province (grant 2020RWG014). The authors gratefully acknowledge all anonymous commentators and Editage for their feedback.

XSL•FO
RenderX

Authors' Contributions

ZF designed, implemented, and wrote the manuscript. HZ, DW, and CF helped to collect and analyze the data. ZC, JH, and DM helped design and implement the study. HG and WY are co-corresponding authors of this study; they read and critically revised the manuscript and contributed to the concept of the research. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. Coronavirus Disease 2019 (COVID-19): Situation Report – 68. Geneva, Switzerland: World Health Organization; 2020 Mar 28. URL: http://apps.who.int/iris/bitstream/handle/10665/331614/nCoVsitrep28Mar2020-eng.pdf?sequence=1&isAllowed=y [accessed 2020-06-04]
3. Wang PW, Lu WH, Ko NY, Chen YL, Li DJ, Chang YP, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: Facebook survey study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
4. Guo B. Comparison of news media under the background of epidemic situation. Journal Res Her 2020;11(13):58-59. [doi: 10.3969/j.issn.1674-8883.2020.13.032]
5. Divecha Z, Divney A, Ickovics J, Kershaw T. Tweeting about testing: Do low-income, parenting adolescents and young adults use new media technologies to communicate about sexual health? Perspect Sex Reprod Health 2012 Sep;44(3):176-183 [FREE Full text] [doi: 10.1363/4417612] [Medline: 22958662]
6. Shaw RJ, Johnson CM. Health information seeking and social media use on the internet among people with diabetes. Online J Public Health Inform 2011;3(1):1-9 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
7. Bao H, Cao B, Xiong Y, Tang W. Digital media's role in the COVID-19 pandemic. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e20156 [FREE Full text] [doi: 10.2196/20156] [Medline: 32530817]
8. Innovations in Singapore's COVID-19 response. UNDP, Singapore Global Centre for Technology, Innovation and Sustainable Development. URL: https://sgtechcentre.undp.org/content/sgtechcentre/en/home/blogs/covid19response.html [accessed 2020-11-04]
9. O'Kane C. Catchy PSA about coronavirus turns into viral TikTok challenge about washing your hands. CBS News. 2020 Mar 04. URL: http://cbsnews.com/news/coronavirus-prevention-vietnam-song-ghen-co-vy-tik-tok-challenge-about-washing-hands-quang-dang-john-oliver/ [accessed 2020-11-05]
10. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: Infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
11. Bastani P, Bahrami MA. COVID-19 related misinformation on social media: A qualitative study from Iran. JMIR Preprints. Preprint posted online on March 27, 2020. [FREE Full text] [doi: 10.2196/18932] [Medline: 32250961]
12. Sriwijitalai W, Wiwanitkit V. Exaggerated information and COVID-19 outbreak. Eur J Clin Invest 2020 May;50(5):e13226 [FREE Full text] [doi: 10.1111/eci.13226] [Medline: 32294236]
13. Ioannidis JPA. Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. Eur J Clin Invest 2020 Apr;50(4):e13222 [FREE Full text] [doi: 10.1111/eci.13222] [Medline: 32191341]
14. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: The case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
15. Fan KS, Ghani SA, Machairas N, Lenti L, Fan KH, Richardson D, et al. COVID-19 prevention and treatment information on the internet: A systematic analysis and quality assessment. BMJ Open 2020 Sep 10;10(9):e040487 [FREE Full text] [doi: 10.1136/bmjopen-2020-040487] [Medline: 32912996]
16. The Lancet. COVID-19: Fighting panic with information. Lancet 2020 Feb 22;395(10224):537 [FREE Full text] [doi: 10.1016/S0140-6736(20)30379-2] [Medline: 32087777]
17. Shen L, Yao R, Zhang W, Evans R, Cao G, Zhang Z. Emotional attitudes of Chinese citizens on social distancing during the COVID-19 outbreak: Analysis of social media data. JMIR Med Inform 2021 Mar 16;9(3):e27079 [FREE Full text] [doi: 10.2196/27079] [Medline: 33724200]
18. Wu P, Zhao D. Research on the identification of key nodes in the process of WeChat epidemic information dissemination: A supernetwork perspective. Math Probl Eng 2020 Aug 27;2020:1-10. [doi: 10.1155/2020/6751686]
19. Zhang L. Research on mobile health education based on WeChat official account. Nurs Rehabil J 2018;17(08):84-86. [doi: 10.3969/j.issn.1671-9875.2018.08.028]

XSL•FO
RenderX

20. Zhang X, Wen D, Liang J, Lei J. How the public uses social media WeChat to obtain health information in China: A survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
21. Company profile. Beijing Qingbo Intelligent Technology. URL: http://home.gsdata.cn/about-us.html [accessed 2020-10-15]
22. WeChat Communication Index WCI (V14.2). Beijing Qingbo Intelligent Technology Co. URL: http://www.gsdata.cn/site/usage [accessed 2020-06-22]
23. Tang J. Research on the Dissemination Effect of Articles Recommended by Guokr WeChat Subscription [master's thesis]. Dalian, China: Dalian University of Technology; 2018. URL: https://d.wanfangdata.com.cn/thesis/ChJUaGVzaXNOZXdTMjAyMTA1MTkSCUQwMTQ3NjczORoIcjM3MzVreHU%3D [accessed 2021-07-13]
24. People's Daily WeChat official account. WeChat Games. URL: http://www.anyv.net/index.php/viewnews-40635/ [accessed 2020-08-30]
25. People's Daily ranked first in WeChat official accounts. Beijing Qingbo Intelligent Technology Co. URL: http://www.gsdata.cn/rank/wxdetail?wxname=cQmB1DySYJnqdi4n [accessed 2020-08-30]
26. At last! All 31 provinces, autonomous regions and municipalities across the country will all lift the first-level response!. CCTV News. 2020 May 01. URL: http://mp.weixin.qq.com/s/SN5eMTf-Yzhc9Bs7rKVjww?scene=25#wechat_redirect [accessed 2020-06-08]
27. Garfin DR, Holman EA, Silver RC. Cumulative exposure to prior collective trauma and acute stress responses to the Boston marathon bombings. Psychol Sci 2015 Jun;26(6):675-683. [doi: 10.1177/0956797614561043] [Medline: 25896419]
28. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
29. Gu T, Chu Q, Yu Z, Fa B, Li A, Xu L, et al. History of coronary heart disease increased the mortality rate of patients with COVID-19: A nested case-control study. BMJ Open 2020 Sep 17;10(9):e038976 [FREE Full text] [doi: 10.1136/bmjopen-2020-038976] [Medline: 32948572]
30. Wang X, Hao Y, Wu Q, Xu W, Wang X, Dai Y, et al. Visual analysis of official microblogging network public opinion on dengue events. Chin J Prev Med 2018;19(06):401-406. [doi: 10.16506/j.1009-6639.2018.06.001]
31. Liu L. Epidemic report of novel coronavirus pneumonia in frame view: Based on textual structure and content mining of WeChat hot text. News Lovers 2020 May 20;2020(05):17-21. [doi: 10.16017/j.cnki.xwahz.2020.05.005]
32. Yang M, Xiao J, Cai H. Multicollinearity in multivariate analysis and its treatment. Chin J Health Stat 2012;29(04):620-624. [doi: 10.3969/j.issn.1002-3674.2012.04.057]
33. Wang ZH. Analyze the power of news dissemination in epidemic prevention and control. Media Forum 2020;3(17):156-158. [doi: 10.3969/j.issn.2096-5079.2020.17.105]
34. Shen Y. Analysis on the mode and thinking of news dissemination in major public emergencies: Taking the new coronavirus pneumonia epidemic as an example. Rep Notes 2020 Apr 30;2020(11):114-116. [doi: 10.3969/j.issn.1004-3799(s).2020.11.093]
35. Bai T. Novel coronavirus pneumonia transmission strategy analysis during the period of new crown pneumonia in China: Take the WeChat official account of "People's Daily" as an example. J News Res 2020 Mar 25;11(06):16-18. [doi: 10.3969/j.issn.1674-8883.2020.06.006]
36. Lachlan KA, Spence PR, Lin X, Najarian K, Del Greco M. Social media and crisis management: CERC, search strategies, and Twitter content. Comput Human Behav 2016 Jan;54(3):647-652. [doi: 10.1016/j.chb.2015.05.027]
37. Du Y, Sun H. Some thoughts on the implementing of prevention and control policy of COVID-19. Chin J Med Manag Sci 2020;10(03):8-12. [doi: 10.3969/j.issn.2095-7432.2020.03.002]
38. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (text message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: Pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]
39. Liu J, Liu Y. News' guidance of group consciousness in public health emergencies: Taking new coronavirus pneumonia news reports as an example. China Radio TV Acad J 2020 Mar 01;2020(03):26-28, 55.
40. Guo Y, Zhang ZH. Public opinion response and information governance under major epidemics: Taking China's new coronavirus pneumonia epidemic in 2020 as an example. J News Res 2020;11(14):11-14.
41. Wang H, Ji W. Main media's liability and responsibilities in news communication of COVID-19. J Anshan Normal Univ 2020;22(03):97-99, 108.

Abbreviations

WCI: WeChat Communication Index

XSL•FO
RenderX

Edited by C Basch; submitted 06.03.21; peer-reviewed by R Shahriarirad, M Pandin; comments to author 09.05.21; revised version received 28.05.21; accepted 10.06.21; published 16.07.21.

Please cite as:
Fan Z, Yin W, Zhang H, Wang D, Fan C, Chen Z, Hu J, Ma D, Guo H
COVID-19 Information Dissemination Using the WeChat Communication Index: Retrospective Analysis Study
J Med Internet Res 2021;23(7):e28563
URL: https://www.jmir.org/2021/7/e28563
doi:10.2196/28563
PMID:34129515

©Zina Fan, Wenqiang Yin, Han Zhang, Dandan Wang, Chengxin Fan, Zhongming Chen, Jinwei Hu, Dongping Ma, Hongwei Guo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Interactive Platform for COVID-19 Literature Visual Analytics: Platform Development Study

Addy Moran[1], BSc; Shawn Hampton[1], BSc; Scott Dowson[1], BSc; John Dagdelen[2,3], BSc; Amalie Trewartha[2], BSc, MPhil, PhD; Gerbrand Ceder[2,3], BSc, MSc, PhD; Kristin Persson[2,3], BSc, MSc, PhD; Elise Saxon[1], BA; Andrew Barker[1], MSc; Lauren Charles[1], BA, MSc, DVM, PhD; Bobbie-Jo Webb-Robertson[1], BA, MEng, DPhil

[1]Pacific Northwest National Laboratory, Richland, WA, United States

[2]Lawrence Berkeley National Laboratory, Berkeley, CA, United States

[3]Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author:
Bobbie-Jo Webb-Robertson, BA, MEng, DPhil
Pacific Northwest National Laboratory
902 Battelle Blvd, J4-18
Richland, WA, 99352
United States
Phone: 1 5093752292
Email: bj@pnnl.gov

Abstract

Background: Papers on COVID-19 are being published at a high rate and concern many different topics. Innovative tools are needed to aid researchers to find patterns in this vast amount of literature to identify subsets of interest in an automated fashion.

Objective: We present a new online software resource with a friendly user interface that allows users to query and interact with visual representations of relationships between publications.

Methods: We publicly released an application called PLATIPUS (Publication Literature Analysis and Text Interaction Platform for User Studies) that allows researchers to interact with literature supplied by COVIDScholar via a visual analytics platform. This tool contains standard filtering capabilities based on authors, journals, high-level categories, and various research-specific details via natural language processing and dozens of customizable visualizations that dynamically update from a researcher's query.

Results: PLATIPUS is available online and currently links to over 100,000 publications and is still growing. This application has the potential to transform how COVID-19 researchers use public literature to enable their research.

Conclusions: The PLATIPUS application provides the end user with a variety of ways to search, filter, and visualize over 100,00 COVID-19 publications.

(J Med Internet Res 2021;23(7):e26995) doi:10.2196/26995

KEYWORDS

COVID-19; visual analytics; natural language processing; scientific literature; software; online platform; literature; interactive; publish; research; tool; pattern; usability

Introduction

COVID-19 has generated a multitude of challenges for scientific and medical researchers, but one of the unexpected challenges was the pace at which scientific literature emerged. In addition to the continually growing body of research that includes many thousands of publications in a single week, there is also related research on other coronaviruses or comorbidities of interest [1,2]. Computational researchers have been working diligently to assemble this information into minable collections such as CORD-19 [3], CovidScholar [4,5], and LitCovid [6,7]. These data sets are of high value but have limited interaction capabilities. Currently, the primary approach for the scientific community to work with these extremely large corpuses of literature has been through data science–based solutions via search engines and tools that categorize data into facets, which works well for very targeted queries [8,9].

XSL•FO
RenderX

With the onslaught of publications being released to help combat COVID-19, there are multiple solutions to search for information within COVID-19 publications. Examples include the Centers for Disease Control and Prevention's (CDC) COVID-19 PubMed Search Alert [10], where the user can specify certain criteria and, when a new publication gets released that matches the user's conditions, the user gets notified. PubMed Search Alert does not provide any support for viewing or searching currently available publications. The CDC also has the PubMed Clinical Queries [11] that allows search by keywords and filter by category, but there are no visualization capabilities, and it returns a simple list of publications. Data-driven visualizations derived from the contents and metadata of these publications can help guide researchers by distilling down the number of publications into a manageable amount while preserving the theme of the query. A newly released tool CoronaCentral [12] offers an improved interface with some visualizations to make searches simpler through a detailed categorization scheme and offers some basic graphics of data summaries based on these categories. The CovidScholar database also helps users with parsing the data via specific tagging classifications and offers a visualization of word embeddings of subsets of papers [4,5]. However, advanced visual analytics of this expanding corpus requires new data science and software solutions. We present a novel platform PLATIPUS (Publication Literature Analysis and Text Interaction Platform for User Studies), which builds on the comprehensive CovidScholar data set and uses visual analytics to give basic and medical researchers a more user-friendly approach to explore their queries of interest. PLATIPUS is publicly available at [13].

Methods

Data

The literature presented in PLATIPUS is collected from original publishers in collaboration with the COVIDScholar project at the University of California, Berkeley/Lawrence Berkeley National Laboratory [5]. Articles in COVIDScholar are sourced by a system of dedicated web scrapers, document parsers, databases, and machine learning models that process papers and metadata into a standardized format that is amenable for text mining. The data in COVIDScholar includes a culmination of 19 sources, presented in Textbox 1, and consists of academic preprints, peer-reviewed research papers, book chapters, patents, clinical trial descriptions, and data sets, all of which have been made openly available by the original publishers to advance COVID-19 research. COVIDScholar updates their data multiple times per day and PLATIPUS queries the COVIDScholar database and reingests new articles once a day.

Textbox 1. Main sources of data in COVIDScholar collection.

Preprints and non–peer-reviewed articles

- medRxiv
- bioRxiv
- Preprints.org
- PsyArXiv
- Social Science Research Network
- SocArXiv
- ChemRxiv
- National Bureau of Economic Research

Peer-reviewed journal articles

- Elsevier
- PubMed
- CORD-19
- Dimensions

Book chapters

- CORD-19

Patents

- The Lens

Clinical trials

- Dimensions

Data sets

- Dimensions

XSL•FO
RenderX

Text Analytics

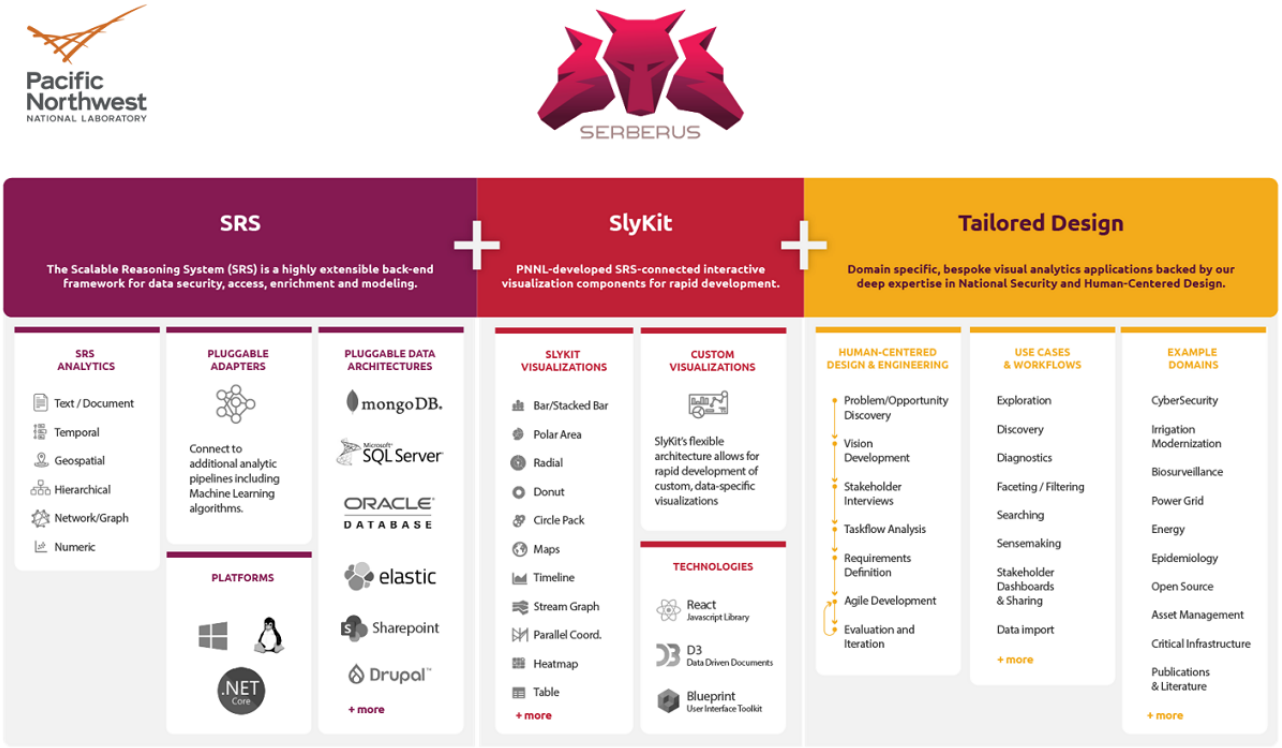
PLATIPUS uses a tool called Automated Analytics and Integration of Data (AAID) to assist in the data ingestion and advanced analytic processing of the COVIDScholar data set. AAID uses multiple algorithms to identify key sources of information while taking into account how the meaning of words change based on the context [14]. AAID uses natural language processing methodologies, specifically entity recognition, machine learning, and human-in-the-loop, to augment the data with additional queryable tags [15]. In PLATIPUS, this means augmenting the COVIDScholar data set with tags such as locations, organizations, diseases, diagnostics and analysis, countermeasures, species, and additional context. AAID uses the NiFi data ingestion and processing pipeline that contains a variety of natural language processing methods such as time-weighted penalized logistic regression models, recursive regex, binary bag of words models, and recurrent neural network models, which is described in detail in Multimedia Appendix 1 Figure S1. The vectorization of the text was based on a bag of words approach. For the clustering visualizations, a k-means default method was used. The analytic capabilities of the AAID pipeline continue to grow to use transformer deep learning classifiers and implement methods to identify anomalies and abnormal characteristics [16].

As of May 2021, there are 159,797 articles that are parsed into various filters. At the top level are authors (n=564,845), categories (n=7), context (n=41), countermeasures (n=28), diagnostics and assay (n=19), disease (n=265), journal (n=11,412), locations (n=365), tags (n=7), species (n=76), and chemicals (n=175). Authors are associated with the publications, and therefore, there are hundreds. For selection purposes, the authors are sorted in order from the most to least prevalent. There are presently seven core categories (treatment, prevention, mechanism, diagnosis, epidemic forecasting, transmission, and case report). Under context, there are 41 groupings associated with the primary context of the article (eg, disease severity or transmission event). Countermeasures are approaches taken against the disease (eg, treatment, vaccine, or awareness campaign). The diagnostics and assay groupings contain the platforms associated with the article, such as transcriptomics or x-rays. Disease is again a broad category where the most prevalent is a categorization of human or animal disease but other specific associated syndrome or special notes are captured here. Journal, similar to author, is a large group of the virtual location of the publication online. Location is a physical location at which the research or case study is conducted for publication, which are extracted using resources from the National Geospatial-Intelligence Agency and United States Geological Survey [17,18]. There are 76 species, the most prevalent being human, rodents, and swine, and 175 chemicals captured that are associated with the manuscripts.

Application Development

PLATIPUS is built on top of the SERBERUS application, which is an end-to-end software solution that rapidly builds visual analytic web applications (Figure 1). Powered by the Scalable Reasoning System (SRS) [19] on the back end and a flexible user interface toolkit on the front end, and drawing from expertise from a user experience and design team, this system is designed for custom solutions that can be readily constructed to support data exploration, discovery, and understanding.

Figure 1. Description of the SERBERUS application full capability components. PNNL: Pacific Northwest National Laboratory.



The PLATIPUS application provides the end user with a variety of ways to search and filter over 100,000 COVID-19 publications. Since PLATIPUS is built on top of SRS and Slykit, PLATIPUS will continue to evolve and grow with new

XSL•FO
RenderX

visualizations and features as SRS and Slykit advances. As of May 2021, PLATIPUS allows the user to filter on locations, categories, authors, organization, disease, diagnostics and analysis, countermeasures, species, and additional context as well as a timeline. The visualizations that are currently available are circle pack, cluster pack, donut graphs, edge-based graph, line chart, matrix, metrics, paracord, table, text clusters, treemap, and timeline described in Textbox 2. The first 10 of these visualizations are at the center of the dashboard and can be assembled based on user choice (one, two, three, etc) all in the view. The timeline visualization is maintained across the top of the user interface. At any time during the filtering and searching process, the user can access a high-level overview of an individual publication, which includes the abstract, information about the authors, tags and categories, and the journal where it was published as well as a direct link to the full publication. Once the user filters down to a subset of publications of interest, they can export the list of publications as a CSV file.

Textbox 2. PLATIPUS (Publication Literature Analysis and Text Interaction Platform for User Studies) core visualizations.

Circle pack

Relative-sized circles of various metadata fields that supports up to three levels (ie, categories→disease→locations)

Cluster graph

Primary properties are clustered into nodes, which are resized based on connection count.

Donut graphs

Data separated based on various properties in a donut circle view where sizes within the donut are relative to frequency

Edge-based graph

Primary property is connected via nodes from a defined link property, which can be filtered based on the number of connections.

Line chart

Multiline chart customized to property selected, data binning, color, and aggregation

Matrix

A 2D grid that shows the aggregations between two properties

Metrics

High-level summary of the data selected

Paracord

Links properties to find connection between metadata, especially useful to find single unique connections

Table

Read-only table format to sort and limit the items being viewed

Text clusters

Groups keywords to place documents into common clusters

Timeline

Bar graph to display metadata over time

Treemap

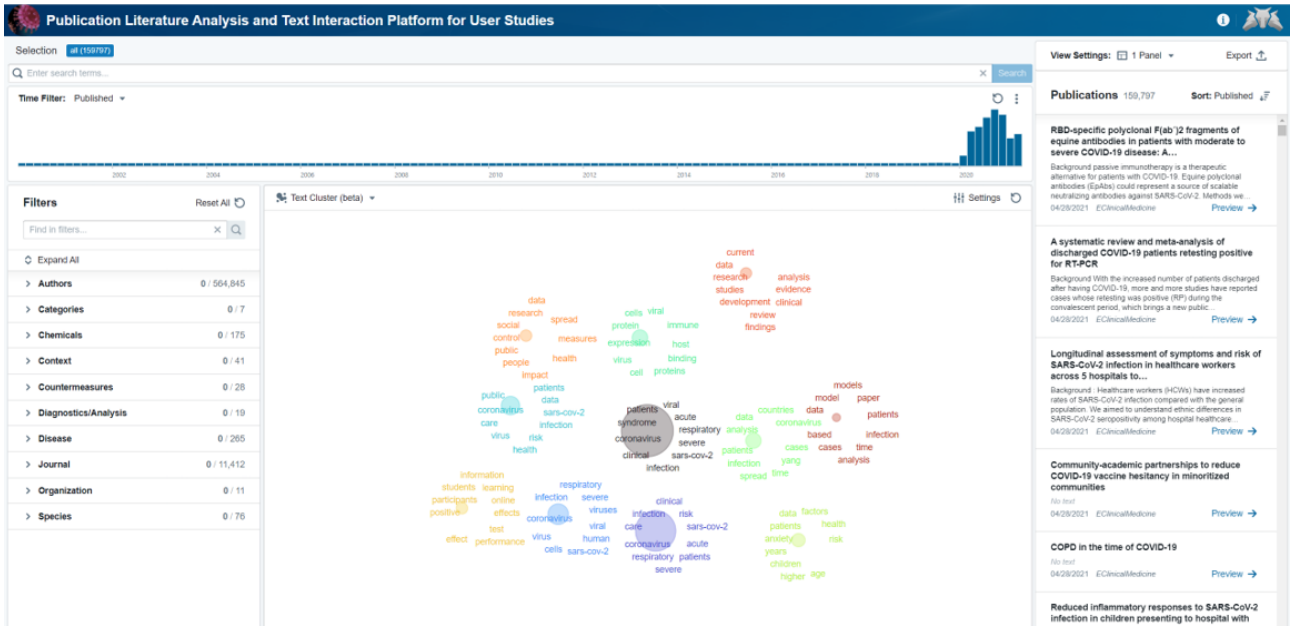
Recursive drill down into subgroups from a primary group

Results

The application allows the user to search by keyword, filter by various tags, select a time range, and visualize the tags and other document properties on innovative graphs and visualizations. Figure 2 shows the home screen of PLATIPUS, which is showing the test clustering view of the full set of COVID-19–related publication literature. PLATIPUS is broken into multiple panels: the search bar on top center, the timeline for filtering articles by date in the center, the filters associated with the annotated data (eg, authors or journals) on the left, the visualization panel (9 total options) in the bottom center, and the article panel (right).

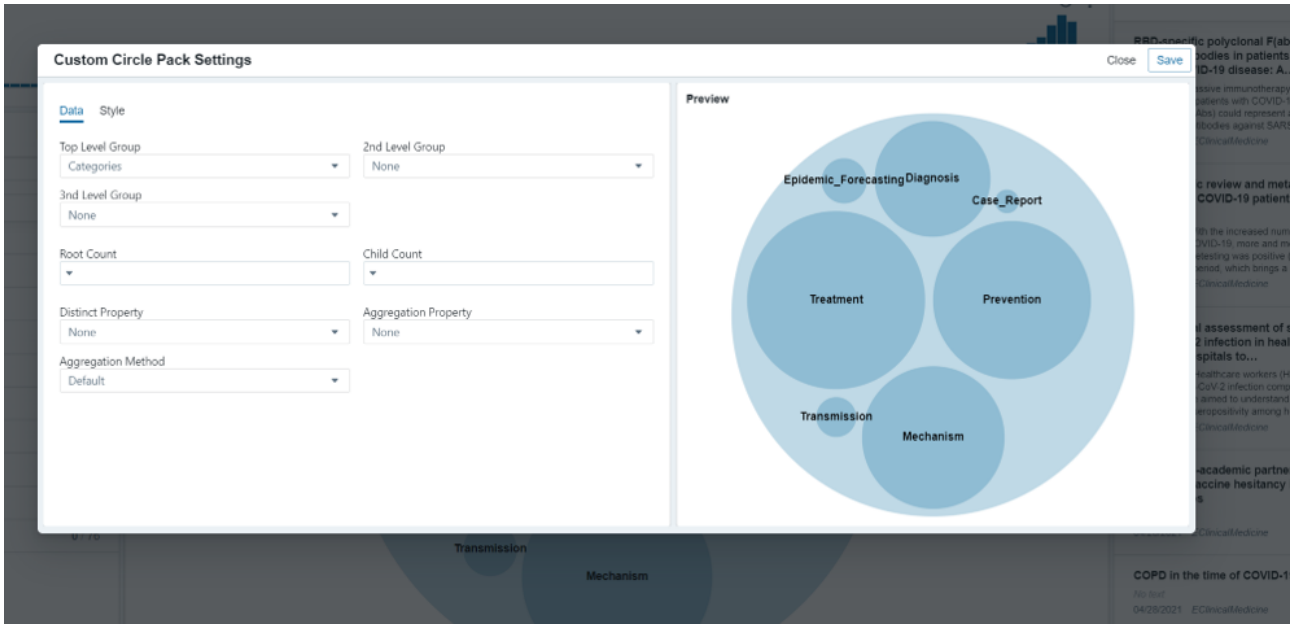
XSL•FO
RenderX

Figure 2. Screenshot of the primary PLATIPUS (Publication Literature Analysis and Text Interaction Platform for User Studies) application page from which the user begins queries and visual interactive activities.



One of the key features of PLATIPUS is the numerous approaches that can be taken to visualize the data. Figure 3 highlights one alternative to the text cluster in Figure 2 (custom circle pack) and how each visualization can be modified to show the specific information of interest to the user. The custom circle pack is driven from the filters on the left-hand side and allows quick views of the overall distribution of this information. For example, for all the COVID-19–related articles in PLATIPUS, we see the majority fall into four core categories: diagnosis, treatment, prevention, and mechanism.

Figure 3. Example of the visualization customization component.

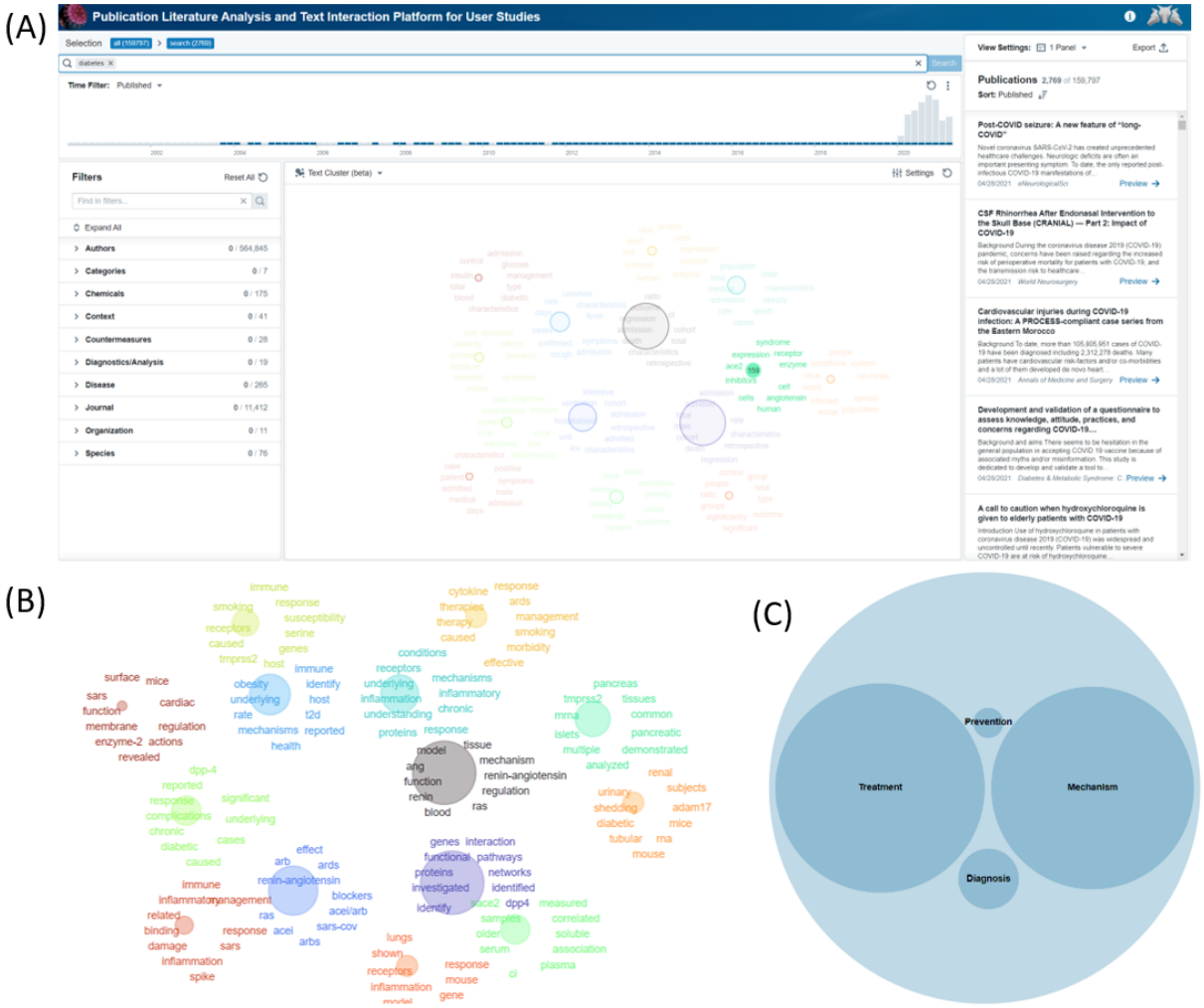


To further explore the functionality of the PLATIPUS application, we demonstrate an example via a case study. There has been significant evaluation of comorbidities such as diabetes on the prognostic response of patients with COVID-19 [20-23]. In this case study, the search of the term "diabetes" in PLATIPUS returns 2769 articles from the originating 159,797, as of May 2021 (Figure 4A). However, this number is too many for a researcher to search through manually. Often the researcher performing the search will select the first few to read in more detail by perusing abstracts or other down-select criteria. This method is still an option within PLATIPUS, as the articles and abstracts are displayed on the right-hand side of the application. A benefit of PLATIPUS is the additional clustering visualization of articles that goes beyond the standard sorting function available in most publication search engines. By evaluating the clusters located in the center of the application (Figure 4A), a researcher interested in the putative receptor angiotensin-converting enzyme 2 can see this is a key cluster in the visualization. Selecting this cluster reduces the literature from 2769 to 159 articles. PLATIPUS then allows the researcher

XSL•FO
RenderX

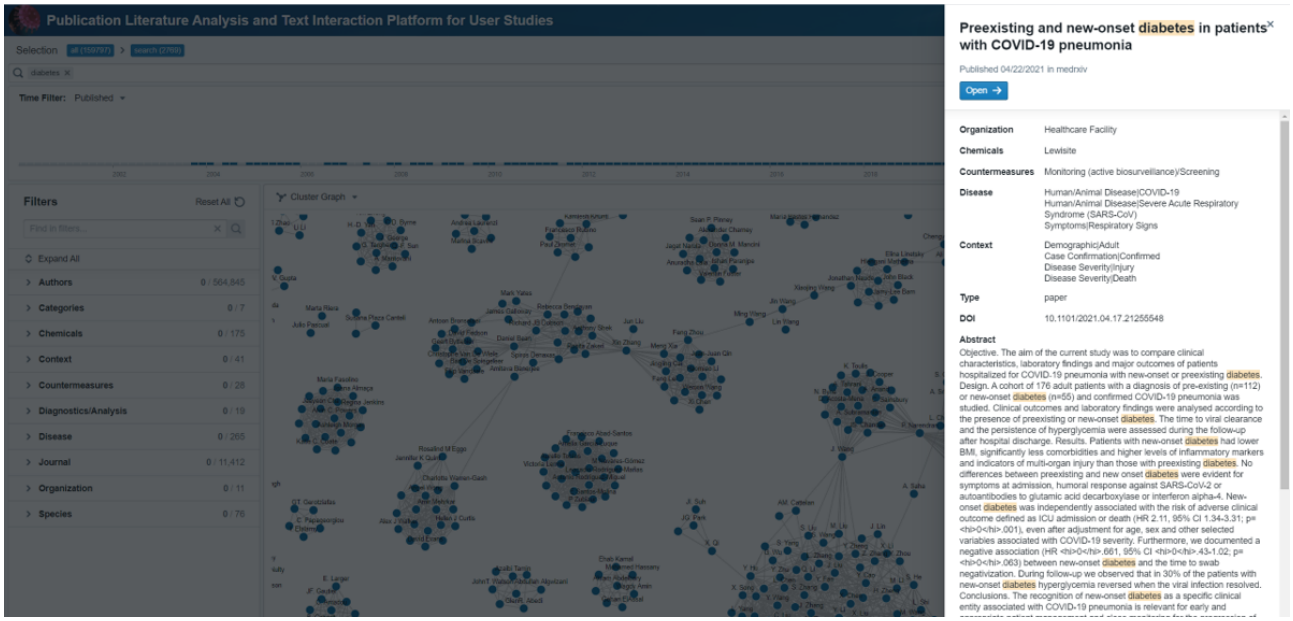
to observe clusters of articles within this new refined query (Figure 4B). The researcher can either narrow down further this way or, as an alternative, can filter articles within the defined facets using a variety of methods (custom circle pack shown in Figure 4C). Within this refined search, the user can view any of the publications via the reading pane. By choosing preview, a publication will open to allow researchers to view the full abstract and associated metadata, and link to the full text, if available, as seen in Figure 5 [24]. Alternatively, on the left side of Figure 4A, there are predefined filters, which include subsets such as “Diagnostics” or “Disease” as an alternate approach to filtering the data. The researcher can also export the metadata from selected documents as a CSV for review in the future.

Figure 4. (A) View and selection based on a search for the term “diabetes” where a cluster associated with angiotensin-converting enzyme 2 (ACE2) is identified, with (B) and (C) as alternate visualizations of the results after selection of the cluster including ACE2.



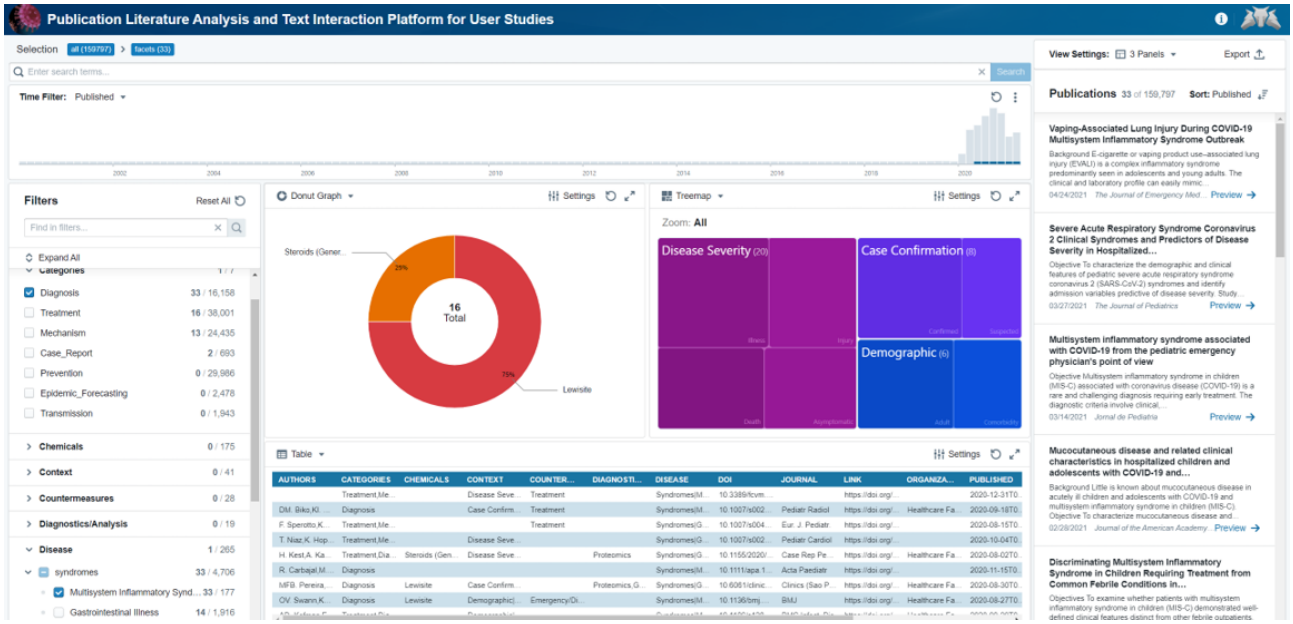
XSL•FO
RenderX

Figure 5. High-level view of a paper, which includes information about COVID-19 and diabetes.



The diabetes example is a visual analytics exploration of a relatively open question, but PLATIPUS also supports direct medical queries using the valuable tagging that is supplied via the AAID pipeline associated with the CovidScholar data. For example, as seen in Figure 6, we applied two filters to find literature that can help with the diagnosis of "Multisystem Inflammatory Syndrome" and "Diagnosis." Multisystem inflammatory syndrome is a new clinical condition due to a cytokine storm associated with COVID-19 that causes inflammation and organ failure [25]. In PLATIPUS, the first filter selected is "Multisystem Inflammatory Syndrome," which reduces the data set to 177 manuscripts. This is further refined into a small set based on the selection of "Diagnosis," which reduces to 33 articles, visible on the left-hand side of Figure 6. The visualizations in this case are tailored to give context of the type of chemical information that is identified from the paper, which may give further insight into how to down-select. The treemap allows the researcher to see the 33 articles that are categorized based on the information of this specific query. Evaluating the 33 articles quickly points to an environmental component of multisystem inflammatory syndrome [26-29].

Figure 6. View and selection based on tagging capability drilling into tags of multisystem inflammatory syndrome and diagnosis.



Discussion

Principal Results

The primary manner the scientific community interacts with scientific literature has, up until recently, not changed in decades. COVID-19 has brought to the forefront of research the challenge of mining literature versus identification of potential articles of interest to a user by keyword searches. To date, PLATIPUS has performed text analytics and clusters, and has visualized nearly 160,000 articles related to COVID-19, and it automatically updates as new documents are added to

XSL•FO
RenderX

COVIDScholar. The application uses state-of-the-art natural language processing (AAID) to provide insight and unique ways to filter and understand the data. PLATIPUS aims to decrease time spent looking through pages of articles by providing the user with multiple ways to search, filter, and view the data. The PLATIPUS application focuses on taking the large amount of literature related to COVID-19 and displaying keywords, categories, and other metadata to allow a user to quickly find relevant information captured by COVIDScholar.

Limitations

PLATIPUS was designed to assist in searching a multitude of COVID-19 publications efficiently, so the user can either find their answer using the visualizations, searching, and drill down capabilities or find a document that will assist in their search. Therefore, PLATIPUS does not support saving views or searches, as it was designed to be a visual analytics search engine and visual table of contents. Additional limitations include the suggestion of the *optimal* visualization based on a query. PLATIPUS allows the users to toggle through visualizations and select those that are of the most utility. Additions to PLATIPUS in the future may be a more guided visualization experience based on the size and complexity of the literature returned from a query. As of March 2021, PLATIPUS does not support finding similar articles to a single selection, but we expect this feature will be available in the future.

Acknowledgments

This research was supported by the Department of Energy (DOE) Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the CARES (Coronavirus Aid, Relief, and Economic Security) Act to the COVID-19 Testing R&D Project. The software was developed at the Pacific Northwest National Laboratory operated by Battelle under contract DE-AC05-76RLO01830. COVIDScholar, an artificial intelligence–powered rapid data gathering, analysis, and dissemination tool, was developed at Berkeley National Laboratory. Work at the Lawrence Berkeley National Laboratory was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under US DOE Contract No. DE-AC02-05CH11231.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PLATIPUS NiFi pipeline diagram.
[PDF File (Adobe PDF File), 28 KB - jmir_v23i7e26995_app1.pdf]

References

1. Editorial. Science in the time of COVID-19. Nat Struct Mol Biol 2020 Apr;27(4):307. [doi: 10.1038/s41594-020-0423-7] [Medline: 32269356]
2. Brainard J. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? Science. 2020 May 14. URL: https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat [accessed 2021-03-17]
3. CORD-19: COVID-19 Open Research Dataset. Semantic Scholar. 2020. URL: https://www.semanticscholar.org/cord19 [accessed 2021-03-17]
4. Trewartha A, Dagdelen J, Huo H, Cruse K, Wang Z, He T, et al. COVIDScholar: an automated COVID-19 research aggregation and analysis platform. arXiv. Preprint posted online on December 7, 2020.
5. Ceder G, Persson K, Dagdelen J, Trewartha A, Huo H, Cruse K, et al. CovidScholar. 2020. URL: https://covidscholar.org/ [accessed 2021-03-17]
6. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature 2020 Mar;579(7798):193. [doi: 10.1038/d41586-020-00694-1] [Medline: 32157233]
7. Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. Nucleic Acids Res 2021 Jan 08;49(D1):D1534-D1540 [FREE Full text] [doi: 10.1093/nar/gkaa952] [Medline: 33166392]
8. COVID-19 Global literature on coronavirus disease. World Health Organization. 2020. URL: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/ [accessed 2021-03-17]
9. COVID-19 portfolio. iCite: NIH Office of Portfolio Analysis. 2020. URL: https://icite.od.nih.gov/covid19/search/ [accessed 2021-03-17]
10. COVID-19 PubMed Search Alert. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/library/researchguides/2019novelcoronavirus/pubmedsearchalert.html [accessed 2021-03-17]
11. PubMed clinical queries. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/clinical/ [accessed 2020-11-03]
12. Lever J, Altman RB. Analyzing the vast coronavirus literature with CoronaCentral. bioRxiv. Preprint posted online on December 22, 2020. [doi: 10.1101/2020.12.21.423860] [Medline: 33398279]

XSL•FO
RenderX

13. PLATIPUS. URL: https://vcs.pnnl.gov/explorer [accessed 2021-06-16]
14. Fellet M. Tracking a pandemic—through words. Pacific Northwest National Laboratory. 2020. URL: https://www.pnnl.gov/news-media/tracking-pandemic-through-words [accessed 2020-11-03]
15. Charles LE, Smith W, Rounds J, Mendoza J. Text-based analytics for biosurveillance. In: Giabbanelli PJ, Mago VK, Papageorgiou EI, editors. Advanced Data Analytics in Health. Cham: Springer; 2018:117-131.
16. Pazdernik K, Charles L. IMS members' work on COVID-19. Institute of Mathematical Statistics. 2020. URL: https://imstat.org/2020/05/17/ims-members-work-on-covid-19/ [accessed 2020-11-03]
17. Complete files of geographic names for geopolitical areas from GNS. National Geospatial-Intelligence Agency. URL: https://geonames.nga.mil/gns/html/namefiles.html [accessed 2021-06-16]
18. U.S. Board on Geographic Names. US Geological Survey. URL: https://www.usgs.gov/core-science-systems/ngp/board-on-geographic-names [accessed 2021-06-16]
19. Scalable Reasoning System. Pacific Northwest National Laboratory. 2020. URL: https://availabletechnologies.pnnl.gov/technology.asp?id=348 [accessed 2021-03-17]
20. Albulescu R, Dima SO, Florea IR, Lixandru D, Serban AM, Aspritoiu VM, et al. COVID-19 and diabetes mellitus: unraveling the hypotheses that worsen the prognosis (review). Exp Ther Med 2020 Dec;20(6):194 [FREE Full text] [doi: 10.3892/etm.2020.9324] [Medline: 33101484]
21. Dennis JM, Mateen BA, Sonabend R, Thomas NJ, Patel KA, Hattersley AT, et al. Type 2 diabetes and COVID-19-related mortality in the critical care setting: a national cohort study in England, March-July 2020. Diabetes Care 2021 Jan;44(1):50-57 [FREE Full text] [doi: 10.2337/dc20-1444] [Medline: 33097559]
22. Kloc M, Ghobrial RM, Lewicki S, Kubiak JZ. Macrophages in diabetes mellitus (DM) and COVID-19: do they trigger DM? J Diabetes Metab Disord 2020 Oct 17:1-4. [doi: 10.1007/s40200-020-00665-3] [Medline: 33102261]
23. Sathish T, de Mello GT, Cao Y. Is newly diagnosed diabetes a stronger risk factor than pre-existing diabetes for COVID-19 severity? J Diabetes 2021 Feb;13(2):177-178. [doi: 10.1111/1753-0407.13125] [Medline: 33107708]
24. Yang J, Lin S, Ji X, Guo L. Binding of SARS coronavirus to its receptor damages islets and causes acute diabetes. Acta Diabetol 2010 Sep;47(3):193-199 [FREE Full text] [doi: 10.1007/s00592-009-0109-4] [Medline: 19333547]
25. Alkan G, Sert A, Oz SKT, Emiroglu M, Yılmaz R. Clinical features and outcome of MIS-C patients: an experience from Central Anatolia. Clin Rheumatol 2021 May 06:1-11 [FREE Full text] [doi: 10.1007/s10067-021-05754-z] [Medline: 33956250]
26. Abrams JY, Godfred-Cato SE, Oster ME, Chow EJ, Koumans EH, Bryant B, et al. Multisystem inflammatory syndrome in children associated with severe acute respiratory syndrome coronavirus 2: a systematic review. J Pediatr 2020 Aug 05:45-54.e1 [FREE Full text] [doi: 10.1016/j.jpeds.2020.08.003] [Medline: 32768466]
27. Carlin RF, Fischer AM, Pitkowsky Z, Abel D, Sewell TB, Landau EG, et al. Discriminating multisystem inflammatory syndrome in children requiring treatment from common febrile conditions in outpatient settings. J Pediatr 2021 Feb;229:26-32.e2 [FREE Full text] [doi: 10.1016/j.jpeds.2020.10.013] [Medline: 33065115]
28. Fernandes DM, Oliveira CR, Guerguis S, Eisenberg R, Choi J, Kim M, Tri-State Pediatric COVID-19 Research Consortium. Severe acute respiratory syndrome coronavirus 2 clinical syndromes and predictors of disease severity in hospitalized children and youth. J Pediatr 2021 Mar;230:23-31.e10. [doi: 10.1016/j.jpeds.2020.11.016] [Medline: 33197493]
29. Hassoun A, Brady K, Arefi R, Trifonova I, Tsirilakis K. Vaping-associated lung injury during COVID-19 multisystem inflammatory syndrome outbreak. J Emerg Med 2021 Apr;60(4):524-530 [FREE Full text] [doi: 10.1016/j.jemermed.2020.12.005] [Medline: 33483200]

Abbreviations

AAID: Automated Analytics and Integration of Data
CARES: Coronavirus Aid, Relief, and Economic Security
CDC: Centers for Disease Control and Prevention
DOE: Department of Energy
PLATIPUS: Publication Literature Analysis and Text Interaction Platform for User Studies
SRS: Scalable Reasoning System

XSL•FO
RenderX

Edited by C Basch; submitted 06.01.21; peer-reviewed by J Soriano, Q Chen, AR Feizi Derakhshi, A Alasmari, A Wahbeh, R Subramaniyam; comments to author 01.03.21; revised version received 29.03.21; accepted 12.06.21; published 16.07.21.

Please cite as:
Moran A, Hampton S, Dowson S, Dagdelen J, Trewartha A, Ceder G, Persson K, Saxon E, Barker A, Charles L, Webb-Robertson BJ
Online Interactive Platform for COVID-19 Literature Visual Analytics: Platform Development Study
J Med Internet Res 2021;23(7):e26995
URL: https://www.jmir.org/2021/7/e26995
doi:10.2196/26995
PMID:34138726

©Addy Moran, Shawn Hampton, Scott Dowson, John Dagdelen, Amalie Trewartha, Gerbrand Ceder, Kristin Persson, Elise Saxon, Andrew Barker, Lauren Charles, Bobbie-Jo Webb-Robertson. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of a Social Media Campaign in Saskatchewan to Promote Healthy Eating During the COVID-19 Pandemic: Social Media Analysis and Qualitative Interview Study

Jordyn L Grantham[1], CRGS; Carrie L Verishagen[1], BSc, RD; Susan J Whiting[1], PhD; Carol J Henry[1], PhD; Jessica R L Lieffers[1], RD, PhD

College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:
Jessica R L Lieffers, RD, PhD
College of Pharmacy and Nutrition
University of Saskatchewan
107 Wiggins Road
Saskatoon, SK, S7N 5E5
Canada
Phone: 1 306 966 8594
Email: jrl210@usask.ca

Abstract

Background: The beginning of the COVID-19 pandemic presented many sudden challenges regarding food, including grocery shopping changes (eg, reduced store hours, capacity restrictions, and empty store shelves due to food hoarding), restaurant closures, the need to cook more at home, and closures of food access programs. Eat Well Saskatchewan (EWS) implemented a 16-week social media campaign, #eatwellcovid19, led by a dietitian and nutrition student that focused on sharing stories submitted by the Saskatchewan public about how they were eating healthy during the COVID-19 pandemic.

Objective: The goal of this study was to describe the implementation of the #eatwellcovid19 social media campaign and the results from the evaluation of the campaign, which included campaign performance using social media metrics and experiences and perspectives of campaign followers.

Methods: Residents of Saskatchewan, Canada, were invited to submit personal stories and experiences to EWS about how they were eating healthy during the COVID-19 pandemic from April to August 2020. Each week, one to three stories were featured on EWS social media platforms—Facebook, Instagram, and Twitter—along with evidence-based nutrition information to help residents become more resilient to challenges related to food and nutrition experienced during the COVID-19 pandemic. Individuals who submitted stories were entered into a weekly draw for a Can $100 grocery gift card. Social media metrics and semistructured qualitative interviews of campaign followers were used to evaluate the #eatwellcovid19 campaign.

Results: In total, 75 stories were submitted by 74 individuals on a variety of topics (eg, grocery shopping, traditional skills, and gardening), and 42 stories were featured on social media. EWS shared 194 #eatwellcovid19 posts across social media platforms (Facebook: n=100; Instagram: n=55; and Twitter: n=39). On Facebook, #eatawellcovid19 reached 100,571 followers and left 128,818 impressions, resulting in 9575 engagements. On Instagram, the campaign reached 11,310 followers, made 14,145 impressions, and received 823 likes and 15 comments. On Twitter, #eatwellcovid19 made 15,199 impressions and received 424 engagements. Featured story submission posts had the best engagement on Facebook and the most likes and comments on Instagram. The EWS social media pages reported increases in their following during the campaign (Instagram: +30%; Facebook: +14%; and Twitter: +12%). Results from the interviews revealed that there were two types of campaign followers: those who appreciated hearing the stories submitted by followers, as it helped them to feel connected to the community during social isolation, and those who appreciated the evidence-based information.

Conclusions: Numerous stories were submitted to the #eatwellcovid19 social media campaign on various topics. On Instagram and Facebook, posts that featured these stories had the highest engagement. During this campaign, EWS's social media following increased by more than 10% on each platform. The approach used for the #eatwellcovid19 campaign could be considered by others looking to develop health promotion campaigns.

(J Med Internet Res 2021;23(7):e27448) doi:10.2196/27448

XSL•FO
RenderX

KEYWORDS

COVID-19; diet, healthy; nutrition; health promotion; social media; dietitian; Saskatchewan

Introduction

On March 11, 2020, the World Health Organization declared the disease caused by the SARS-CoV-2 coronavirus, COVID-19, a pandemic [1]. COVID-19 is a respiratory disease with many symptoms, including fever, dry cough, and fatigue [2]. Respiratory droplets are the main source of disease transmission, and individuals are most likely to become infected when they are in close contact with others who have COVID-19 [3]. Although approximately 80% of people who develop symptoms from COVID-19 recover without hospitalization, this disease is particularly concerning as about 15% of people who have symptoms of COVID-19 require hospitalization, and about 5% require intensive care treatment [2]. Containment measures, including restrictions on travel; closing nonessential businesses, services, and schools; requiring and encouraging people to work from home; social distancing; handwashing; and use of face masks, have been put in place in attempts to contain the virus [4-7].

The COVID-19 pandemic presented many challenges related to food, nutrition, and eating habits. Food access and disruptions to food environments were a major concern at the beginning of the pandemic. The social distancing measures also caused abrupt changes in food sourcing patterns [8]. One of the first impacts was a stage of food hoarding and panic shopping; in Canada, grocery store sales increased by 38% the week of March 11, 2020, compared to 2019, with the sale of dried goods, frozen foods, and shelf-stable foods having greatly increased [9]. Residents also experienced price fluctuations that limited purchasing power, and there were increases in demand for online grocery shopping [10]. Food hoarding and panic buying behaviors that took place early in the pandemic reflected consumers' concerns regarding decreased food availability and uncertainty in the stability of the supply chain [11]. Grocery shopping habits and patterns were disrupted as consumers attempted to comply with the new public health recommendations, which included advising that a single member of the household shop for food biweekly [10,12,13]. In addition, numerous social distancing measures were put in place at food stores, which further complicated grocery shopping [10]. Purchasing food to last for 2 weeks was overwhelming, especially for those with reduced purchasing power, impacting the types of meals families could prepare and requiring increased planning, a difficult task when food availability is uncertain and time is limited in food stores [10,13]. Prior to the pandemic, there were also concerns surrounding where the public obtains nutrition information [14,15], and the pandemic may have created more concern around this. In addition, restaurants and smaller food retailers were suddenly closed or had to take time to shift service styles, disrupting the food environment further [16,17]. There were also concerns about food safety (eg, handling of produce at the grocery store and transmission via surface contamination of frozen foods) [18-21]. Vulnerable populations, those already food insecure, and those experiencing employment instability were most affected by the abrupt disruptions to food environments and food access issues and were more likely to experience food insecurity as a result, but these food access issues impacted people across socioeconomic levels [22-25].

The abrupt changes to the daily routines due to COVID-19 had impacts on food preparation and eating habits; studies have reported both positive and negative impacts. For example, many people began working from home or participating in online learning, and this led to increased time spent cooking, more time eating together as a family, reduced take-out food consumption, increased overall food consumption, and increased calories from snack food consumption [26]. More time spent at home also meant parents were fielding increased snack requests from children and dealing with the temporary loss of school-provided meals [22,26,27]. Disruptions to school food programs impacted vulnerable populations and food insecure households disproportionately. However, some studies have reported improvements in diet quality and eating habits during the early pandemic [13,28]. In addition, studies have also reported increased sedentary behavior, poorer dietary intakes, and increased alcohol consumption during the COVID-19 pandemic among university students [29].

There has also been a substantial amount of misinformation shared regarding COVID-19, including misinformation related to food and nutrition shared on social media [30-33]. The link between nutrition and immune health is important, and a well-rounded and varied diet supports immune health, but there was an increase in unsubstantiated claims that certain eating patterns, diets, foods, products, or supplements can prevent or cure COVID-19 [31,34-36]. Based on the sheer volume of sudden and novel challenges presented to the public, public health professionals had to quickly adapt services to adhere to social distancing measures and develop strategies to assist the public in dealing with these new challenges, including dealing with misinformation. Van den Broucke [32] suggested that health promotion strategies that promote community engagement can improve health outcomes and strengthen community capacity to address pandemic-related disruptions, and this was supported by others in respect to the COVID-19 pandemic [33,37].

One option that warrants further attention in this situation is social media, as it allows individuals to connect and support one another virtually and facilitate community engagement, and it has been gaining popularity in the dietetic profession and for public health use [38,39]. Social media has also been found to be a cost-effective tool to disseminate health promotion messages across populations that may not be reached using traditional advertising methods, such as minorities or lower socioeconomic groups, and can aid in building a supportive virtual community [40]. In Canada, social media use is very common; a recent report found that 94% of Canadian adults who use the internet have an account on at least one social media platform, and 83%, 51%, and 42% of these individuals had a Facebook, Instagram, and Twitter account in 2020, respectively

XSL•FO
RenderX

[41]. This same report also found that in Canada, Instagram had the largest gain among adults who had an account (+14% since 2017) [41]. In addition, during the COVID-19 pandemic, use and analysis of social media has generated substantial interest [42-46]. For example, Pahayahay and Khalili-Mahani [42] investigated the relationship between stress and the media during the first 4 weeks of lockdown due to COVID-19 and found that respondents had a complex relationship with social media. They found that they rely on it for a feeling of social connection and as an information source during isolation. However, they also found that the increased exposure to negative news or misinformation surrounding COVID-19 caused increased stress [42]. Pahayahay and Khalili-Mahani [42] further found that there was a need for positivity in media (eg, news social networks and newsletters) during this time.

Eat Well Saskatchewan (EWS) [47] is a free service that connects residents of the province of Saskatchewan, Canada, to a registered dietitian and provides evidence-based food and nutrition information by phone, email, and social media. EWS noticed an increase in social media use in late March 2020 and, as a result, created evidence-based content related to nutrition and COVID-19 and posted a COVID-19 and nutrition FAQ (frequently asked questions) section on their website [48] to answer questions surrounding immune health, grocery shopping, breastfeeding, and food safety. EWS also began disseminating COVID-19–related social media posts early in the pandemic and saw a 20% spike in performance of posts, suggesting that there was public interest in receiving evidence-based nutrition information from a reliable source. Because of these observations, EWS implemented a positive social media campaign, #eatwellcovid19, that invited residents to share personal stories about how they were eating healthy during the COVID-19 pandemic.

This manuscript seeks to (1) describe the implementation of the #eatwellcovid19 social media campaign and (2) share results from the evaluation of the campaign, which encompassed examining campaign performance using social media metrics and experiences and perspectives of campaign followers using semistructured interviews.

Methods

Ethics Approval

This research project was approved by the University of Saskatchewan Behavioural Research Ethics Board (BEH 1975). The analysis of social media analytics received an exemption from the University of Saskatchewan Behavioural Research Ethics Board.

Setting

Saskatchewan is a multicultural prairie province in Canada that spans over 588,000 square kilometers [49]. The most recent Canadian census was done in 2016 and found that Saskatchewan had a population of 1,098,352, of which 112,490 reported as immigrants; 115,875 reported being a visible minority; and 175,020 reported an Aboriginal identity [49]. Saskatchewan Bureau of Statistics reported that 655,313 residents lived in urban cities; 149,717 lived in towns; 176,535 lived in rural municipalities; 47,308 lived in villages; 56,050 lived on reserve; and 13,429 lived in other areas, such as Crown Colonies or hamlets, in 2016 [50]. The Saskatchewan Bureau of Statistics estimates that as of July 1, 2020, the population of Saskatchewan was 1,178,681 people [51].

Social Media Campaign Design

This social media campaign, #eatwellcovid19, was led by the registered dietitian at EWS and assisted by the EWS student assistant and researcher, a second-year nutrition student; three nutrition faculty members provided guidance and advice for the campaign on an as-needed basis. EWS can be found on Facebook as Eat Well Saskatchewan, on Instagram as @eatwellsaskatchewan, and on Twitter as @EatWellSask.

In the past, EWS has conducted social media campaigns. One previous campaign (ie, #eatwellchampion) provided a platform for Indigenous people from across Saskatchewan to tell their own stories about the impact that nutrition has had on them. An informal evaluation of social media metrics found that #eatwellchampion posts performed well (eg, commonly shared) and appeared to resonate with followers. The #eatwellcovid19 campaign was partially based on the successful experience of this campaign.

The purpose of the #eatwellcovid19 campaign was to provide a platform for Saskatchewan residents to share their experiences and stories about how they are eating well during the COVID-19 pandemic. EWS announced the #eatwellcovid19 campaign on their social media platforms—Facebook (Facebook, Inc), Instagram (Facebook, Inc), and Twitter (Twitter, Inc)—on April 20, 2020, through a call-out advertisement encouraging residents of Saskatchewan to submit stories on social media platforms (ie, #eatwellcovid19, @eatwellsaskatchewan, and @eatwellsask) or by email; a sample digital call-out poster is presented in Figure S1 in Multimedia Appendix 1. Saskatchewan residents were invited to submit personal stories about how they were eating healthy during the COVID-19 pandemic. They were provided with sample topics and a framework for what to submit. Stories could be submitted using videos, pictures, and/or words. The campaign ran for 16 weeks and, during this time, the EWS team promoted the campaign on their social media platforms through community partners, such as Indigenous Services Canada, and through the University of Saskatchewan online bulletin. Social media advertising was also purchased to promote the campaign beyond the organic reach. Advertisements were created to encourage story submissions from the public and different vulnerable populations. A total of Can $180 was used for paid advertising to encourage story submissions. On Facebook, three posts to encourage story submissions were boosted with advertisements: 12 days for Can $50, 20 days for Can $50, and 20 days for Can $50. In addition, Can $30 was paid to encourage story submissions on Instagram for an unknown duration, and no paid advertisements to encourage story submissions were used on Twitter. In addition, to encourage story submissions, there was a weekly draw for a Can $100 grocery gift card.

To appeal to a variety of followers and address different areas of interest, the campaign included a variety of posts, including stories, supplemental content related to stories, information

XSL•FO
RenderX

resources, and winner announcements. Samples of the different types of social media posts are available in Figures S2 to S7 in Multimedia Appendix 1. The EWS team selected story submissions with unique or mass appeal to be developed into feature stories. One to three feature stories were then featured on the EWS social media platforms each week. Supplemental content was curated by the EWS team and was used to complement featured story submissions and provide more information on the subject matter. This included evidence-based nutrition information, links for further reading from reputable sources, recipes, and tips related to stories. For example, if a story sent in by a Saskatchewan resident described eating more berries during the pandemic, nutritional information about berries, recipes, and tips on picking and preparing berries would be shared to complement this story. Of note, no specific criteria were used in choosing the recipes that were shared; a variety of recipes were shared to appeal to the wide demographic of followers. Some of the recipes shared were submitted by campaign followers and others were found and shared by EWS staff. Examples of recipes shared included green pesto rice, watermelon fruit pizza, cattail pollen biscuits, and one for pickled carrots to encourage pickling and to provide education on food preservation. Information resource posts were those that included general COVID-19 nutrition and food information, content shared from other social media pages, and helpful external links not related to any one story. Winner announcements were posts that announced the winner of a gift card for that week. Posters were also used to advertise the campaign and encourage story submissions.

EWS also utilized the 24-hour story feature on Instagram and Facebook to interact with followers and promote the campaign. However, these 24-hour stories should not be confused with the featured story submissions. The content posted using the 24-hour story feature varied and utilized the many interactive features available on the social media applications (eg, quizzes). These features allowed EWS to promote the campaign, conduct polls, provide interactive quizzes, and ask questions to followers.

The EWS team posted #eatwellcovid19 content in a dynamic manner rather than structuring content to be posted at the same time or on a specific day, largely because the content posted was the result of public participation. The EWS team also engaged with EWS social media followers. This engagement was done by promptly responding to and *liking* comments left by followers on posts and responding to and communicating with followers using the social media platforms' direct messaging features as appropriate (eg, responding to a follower's private message).

Social Media Analytics Analysis

Reach, impressions, engagement, and engagement rate regarding the #eatwellcovid19 campaign were determined using Facebook Insights, Instagram Insights, and Twitter Analytics. The definitions of the different social media metrics vary across each social media platform and are provided in Table 1 [52-55].

To begin the social media analytics analysis, the EWS student researcher (JLG) separated #eatwellcovid19 posts from other posts on the EWS pages that were shared during this time. Reach, impressions, and engagements of the #eatwellcovid19 campaign were determined separately for each social media platform, as the metrics provided by each platform are different. Each #eatwellcovid19 post was classified into a category, depending on the type of content shared (eg, *featured story submission* or *campaign poster*). The metrics provided for individual posts were then recorded. Descriptive statistics, specifically the mean and standard deviation of reach, impressions, engagement, and engagement rate for each post category on each platform, were determined using Microsoft Excel (Microsoft Corp). Metrics for Twitter were collected on October 2, 2020, and metrics for Facebook and Instagram were collected on October 5, 2020. Metrics for the 24-hour Instagram and Facebook stories were not captured because of an oversight, as this information disappears soon after the post is made (eg, 28 days on Facebook). Because data collection occurred after this time period, this information was no longer available.

XSL•FO
RenderX

Table 1. Definition of key terms for each social media platform.

Key term	Definition
Facebook	
Reach	The number of unique individuals who viewed the listed type of #eatwellcovid19 post [52]
Engagement	The number of actions that individuals made with the listed type of #eatwellcovid19 post, including likes, comments, shares, post clicks, etc [53]
Impressions	The number of times the listed type of #eatwellcovid19 post entered an individual's screen (may include more than one view from the same individual) [52]
Instagram photos	
Reach	The number of individuals who saw the listed type of #eatwellcovid19 post [54]
Impressions	The number of times the listed type of #eatwellcovid19 post was seen (may include more than one view from the same individual) [54]
Likes	The number of engagements that involved pressing the *like* button on the listed type of #eatwellcovid19 post
Comments	The number of engagements that involved leaving a comment on the listed type of #eatwellcovid19 post
Instagram videos	
Reach	The number of individuals who saw the listed type of #eatwellcovid19 post [54]
Likes	The number of engagements that involved pressing the *like* button on the listed type of #eatwellcovid19 video
Views	The number of times the listed type of #eatwellcovid19 video was viewed
Twitter	
Impressions	The number of times the listed type of #eatwellcovid19 post entered a person's screen (eg, timeline and search results) [55]
Engagement	Any interactions on the listed type of #eatwellcovid19 post, including clicks, retweets, replies, likes, etc [55]
Engagement rate	This value is calculated by Twitter and is total engagements divided by total impressions for the listed type of #eatwellcovid19 post [55]

Semistructured Qualitative Interviews

Reporting of the qualitative interview portion of this manuscript was guided by the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist [56]. Purposeful sampling [57] was used to select participants to complete a qualitative interview on their experiences and perspectives regarding the #eatwellcovid19 social media campaign. Participants were recruited via the EWS social media pages. A poster advertising the study was posted on the EWS social media pages in August and September 2020. One poster to advertise this study was promoted on Facebook with paid advertising (Can $25 for 5 days). Prospective participants contacted the researcher via telephone or email. After the prospective participant contacted EWS, the researcher screened the individual to determine if they met the study inclusion criteria. Inclusion criteria were as follows: Saskatchewan resident, ≥18 years of age, can speak English, had followed EWS social media accounts, and had the ability to provide informed consent. Once ensuring the participant met the inclusion criteria, the researcher scheduled an interview with the participant via telephone or online.

All telephone or online interviews were conducted remotely by the EWS assistant and student researcher (JLG), a female nutrition student trained in qualitative interview methods. The only individuals present during the interviews were the interviewer and the participants. All interviews occurred in September 2020. There were no established relationships between the interviewer and participants. Prior to beginning the interview, the researcher read the consent form to the participant, which detailed the study objectives and procedures, ensured the participant's understanding, and obtained oral consent. The interview protocol (Multimedia Appendix 2) contained questions about EWS follower experiences and perspectives regarding the #eatwellcovid19 social media campaign; both clarifying and elaborating probes were used to gather additional information. The interview protocol was developed by all members of the research team and field-tested with one individual to ensure that the questions were appropriate, relevant, and clear and captured the desired information [58]. Interviews were completed until data saturation occurred. Creswell defines data saturation as when "the researcher stops collecting data because fresh data no longer sparks new insights or reveals new properties" [59]. No repeat interviews were conducted. After completion of the interview, participants were mailed a Can $25 grocery store gift card in appreciation of their time. No field notes were taken during the interviews. All interviews were audio recorded and transcribed verbatim by a professional transcriptionist. Transcripts were not returned to participants for comment or corrections. Analysis of the interview transcripts was conducted using conventional content analysis [60]. All interview transcripts were coded inductively by the student researcher (JLG) using NVivo 12 (QSR International). The coding process allowed inductive categories and themes to emerge from the data [60]; no a priori codes were used. All interview transcripts, coding, and the draft themes were reviewed by a second member of the research team (JRLL)

XSL•FO
RenderX

with experience in qualitative research. Discrepancies were discussed and consensus was reached.

Results

Overview

The #eatwellcovid19 campaign received 75 story submissions from 74 Saskatchewan residents, and 42 of these stories were featured on EWS social media pages. The main themes seen in the submitted stories about how Saskatchewan residents were eating healthy during the COVID-19 pandemic can be found in Table 2. The top three themes of submitted stores were increased time spent cooking or trying new recipes, grocery shopping habits, and gardening. Of note, 2 featured stories appeared in local Saskatchewan newspapers and the top featured story differed by platform.

Table 2. Themes of submitted stories for the #eatwellcovid19 social media campaign.

Story theme	Stories submitted (N=75), n (%)	Examples
Increased time spent cooking or trying new recipes	21 (28)	Budget-friendly recipes, trying different cuisines, bread baking, and improved nutrition from more homemade meals
Grocery shopping habits	17 (23)	Stocking up on nonperishable items, change in frequency of grocery shopping trips, feelings of unease in the grocery store, and individual vs family outing
Gardening	12 (16)	Tried gardening for the first time, expanded garden, and utilized a community garden
Traditional skills	8 (11)	Foraging, hunting, and preserving
Increased family time and family meals	6 (8)	Change in mealtime patterns, more time to eat together, and involving kids in the kitchen
Other	11 (15)	Food waste and pregnancy

Social Media Analytics

EWS's Facebook page began the campaign at 1524 page likes and ended the campaign with 1780 page likes, representing 14% growth. A notable outcome for EWS's Instagram page was the 30% growth in followers from the beginning of the campaign. The EWS Instagram account began the campaign with 262 followers and ended the campaign with 376 followers. The EWS Twitter account had 132 followers at the beginning of the campaign and ended the campaign with 150 followers, representing 12% growth. Analyses of social media metrics for the different types of posts are shown in Tables 3-6.

The focus of the #eatwellcovid19 campaign was on Facebook, as this was where EWS had the largest social media following. In total, 100 #eatwellcovid19 posts, not including 24-hour stories, were shared on the EWS Facebook page and, overall, there were 9575 engagements with this content on Facebook. On Facebook, after excluding post types that were sometimes promoted with paid advertising, featured story posts (n=38) had the highest mean engagement (133 engagements/post). The mean engagement of these posts was more than three times higher than the mean engagement of supplemental content related to featured story posts (40 engagements/post), more than four times higher than the mean engagement for winner announcement posts (30 engagements/post), and more than six times higher than the mean engagement for information resource posts (20 engagements/post). The campaign was also seen by many individuals on Facebook. In total, the #eatwellcovid19 campaign had a reach of 100,571 and left 128,818 impressions. Similar to engagement, for post types that were not promoted with paid advertising, featured story submissions had higher reach and impressions compared to other types of posts.

Table 3. Metrics of the #eatwellcovid19 social media posts on Facebook.

Facebook post category	Posts or stories, n	Engagement, mean (SD)	Reach, mean (SD)	Impressions, mean (SD)
Featured story submission	38	133 (212)	824 (773)	987 (959)
Supplemental content related to featured story	24	40 (46)	576 (684)	673 (767)
Campaign poster[a]	8	361 (487)	5433 (5353)	7679 (7340)
Information resource	23	20 (13)	360 (84)	410 (95)
Winner announcement	7	30 (17)	526 (224)	612 (225)
24-hour story	27[b]	N/A[c]	N/A	N/A

[a]Some campaign posters on Facebook were boosted with paid advertising.

[b]This number is approximate.

[c]N/A: not applicable; this information was unable to be collected from Facebook.

XSL•FO
RenderX

Table 4. Metrics of the #eatwellcovid19 social media posts on Instagram.

Instagram post category	Posts or stories, n	Likes, mean (SD)	Comments, mean (SD)	Reach, mean (SD)	Impressions, mean (SD)
Featured story submission	36	15 (8)	0.4 (0.8)	171 (38)	214 (52)
Supplemental content related to featured story	5	11 (5)	0.2 (0.4)	160 (16)	183 (21)
Campaign poster[a]	5	23 (30)	0 (0)	584 (959)	850 (1494)
Information resource	2	10 (7)	0 (0)	153 (28)	177 (42)
Winner announcement	5	12 (6)	0.2 (0.4)	160 (27)	187 (39)
24-hour story	58	N/A[b]	N/A	N/A	N/A

[a]Some campaign posters on Instagram were boosted with paid advertising.

[b]N/A: not applicable; this information was unable to be collected from Instagram.

Table 5. Metrics of the #eatwellcovid19 social media videos on Instagram.

Instagram video category	Posts, n	Likes, mean (SD)	Reach, mean (SD)	Views, mean (SD)
Featured story submission	2	14 (3)	155 (12)	103 (27)

Table 6. Metrics of the #eatwellcovid19 social media posts on Twitter.

Twitter post category	Posts, n	Engagements, mean (SD)	Engagement rate[a], mean (SD)	Impressions, mean (SD)
Featured story submission	26	10 (10)	3 (2)	281 (212)
Supplemental content related to featured story	3	10 (10)	3 (0.6)	339 (368)
Campaign poster	7	18 (18)	2 (1)	890 (830)
Information resource	1	1 (0)	1 (0)	218 (0)
Winner announcement	2	4 (2)	2 (1)	217 (0.7)

[a]This value is calculated by Twitter and is total engagements divided by total impressions for the listed type of #eatwellcovid19 post [55].

Results on Instagram campaign performance using social media metrics were similar to Facebook. In total, 55 #eatwellcovid19 posts, not including 24-hour stories, were shared on EWS's Instagram page and these posts received a total of 823 likes and 15 comments. After excluding post types that were sometimes promoted with paid advertising, featured story posts (n=36) had higher numbers of mean likes (15 likes/post) and mean comments (0.4 comments/post) compared to other types of posts, such as winner announcements (mean 12 likes/post and mean 0.2 comments/post), supplemental content related to the featured story (mean 11 likes/post and mean 0.2 comments/post), and information resource posts (mean 10 likes/post and mean 0 comments/post). The campaign was also seen by many followers on Instagram. In total, on Instagram, the #eatwellcovid19 campaign had a reach of 11,310 and left 14,145 impressions. Similar to engagement, for post types that were not promoted with paid advertising, featured story submissions had higher reach and more impressions compared to other types of posts.

Results on Twitter campaign performance using social media metrics were different compared to Facebook and Instagram. In total, 39 #eatwellcovid19 posts were shared on the EWS Twitter page and these posts received 424 engagements. Regarding engagement, the campaign posters received the most engagements (mean 18 engagements/post), and this was followed by featured story submissions (mean 10 engagements/post), supplemental content related to featured story posts (mean 10 engagements/post), winner announcements (mean 4 engagements/post), and information resource posts (mean 1 engagement/post). The campaign was also seen by many followers on Twitter. In total, there were 15,199 impressions, averaging an engagement rate of 3%. Impressions were highest for campaign posters, followed by supplemental content related to featured story posts, featured story posts, information resource posts, and winner announcement posts.

Semistructured Qualitative Interviews

In total, 86 people indicated interest in participating in a qualitative interview, 35 people responded to the initial screening to proceed further, 12 were ineligible based on pre-established criteria, and 20 interviews were conducted. No participants dropped out of the study after the interviews took place. The average interview took 8 minutes and 54 seconds (range 4 minutes and 39 seconds to 12 minutes and 31 seconds). Participant demographics are listed in Table 7.

XSL•FO
RenderX

Table 7. Demographics of interview participants.

Demographic	Participants (N=20), n (%)
Age category (years)	
18-30	8 (40)
31-50	10 (50)
51-70	2 (10)
Gender	
Female	18 (90)
Male	2 (10)
Other	0 (0)
Prefer not to say	0 (0)
Place of residence	
Urban	17 (85)
Rural	3 (15)
Remote	0 (0)
Immigrant status	
Canadian citizen or nonrecent immigrant (>5 years residing in Canada)	19 (95)
Recent immigrant (<5 years residing in Canada)	1 (5)

The participants followed EWS for variable amounts of time; of note, most of these participants followed EWS on Facebook and Instagram, and only one participant followed EWS on Twitter. Analysis of the interview transcripts showed that followers appreciated the content shared during the #eatwellcovid19 campaign. There appeared to be two types of campaign followers. The first type of follower valued the positive personal stories shared in a time with increased negative news, the stories resonated with them, and they felt a connection to their community by following this campaign. A notable quote from an interviewee demonstrated that the storytelling format provided a connection to the community in a time where physical connection was not possible:

> *I think that there's a lot of value in taking people's stories, lived experiences, and sharing it with others because I find, a lot of the times, we are living in silos and we're all thinking, "I'm the only one who struggles with food insecurity. I'm the only one that can't figure out how to make healthy meals for my kids that they'll actually eat," and really that's not the case at all, but with the opportunity that comes with working with other people that says, "Me too. I struggle with this as well." It brings connection and community within these silos we have and just makes us know that we're not alone in our struggles. Everybody has concerns about food and how to be healthy.* [Participant #13, female, 18 to 30 years of age]

The second type of follower was content driven; these followers appreciated the evidence-based information given and followed the campaign for recipes, meal planning tips, and food safety information during the COVID-19 pandemic. A notable quote from an interviewee demonstrated that the campaign helped them to overcome grocery shopping challenges during the COVID-19 pandemic:

> *There was this link to "How to grocery shop and prioritize during COVID." And there was lots of information about, "Don't over-buy. Make sure you have this on stock and this on stock," and it's really good for me because I'm kind of not the great—I don't have too much income, and so I have to prioritize and budget pretty well. So seeing a list like that was pretty good for helping me to try to prioritize what I need during this time.* [Participant #14, female, 18 to 30 years of age]

Both types of followers appreciated and emphasized the importance of sharing local, Saskatchewan-based content to meet the unique needs of Saskatchewan residents.

Discussion

Principal Findings

Overall, numerous stories on various topics related to healthy eating were submitted to the #eatwellcovid19 social media campaign. In general, the social media metrics revealed that featured stories had better engagement compared to other types of posts (eg, supplemental content related to featured stories). The #eatwellcovid19 campaign also reached many individuals and made numerous impressions on social media. The semistructured interviews revealed that there were two types of campaign followers: those who appreciated reading the featured story posts, as they found that these posts helped them to feel connected to their community, and those who appreciated the evidence-based information. The lessons learned from #eatwellcovid19 are helpful to those looking to develop their own social media campaigns for health purposes.

XSL•FO
RenderX

In total, EWS received 75 story submissions from 74 individuals for #eatwellcovid19 over the 16-week campaign, which was substantial. The topics of the submitted stories were broad and included, for example, gardening, traditional skills, grocery shopping, family time, and meals. Because of the broad nature of story topics, they seemed to appeal to a wide variety of followers. This campaign was also unique, as much of the campaign content was organic and provided by the social media followers rather than developed by researchers or health professionals, which is more typical in social media campaigns [61]. We also found that Can $100 grocery gift card draws were likely a good incentive for the public to submit their personal stories. This finding is not surprising, as Perrault et al also found in their survey study on college health social media pages that incentives were desired by students to follow, interact with content, and share content with others [62]. Of note, a few of the gift card winners commented on how helpful the Can $100 grocery gift card was going to be in the economically challenging times.

In terms of engagement on Facebook, and likes and comments on Instagram, featured story submissions appeared to perform the strongest relative to other types of campaign posts, including supplemental content related to the featured story, winner announcements, and information resources. This finding is similar to a study by Gabarron et al who found that on the Facebook, Instagram, and Twitter pages for the Norwegian Diabetes Association, posts containing stories and personal interviews had the most likes, comments, and shares compared to other types of posts over a 3-year period [63]. They also found that posts with information on recipes and food-related information received less engagement [63]. In addition, Pedersen et al also found that in a social media campaign on human papillomavirus vaccine promotion, posts with personal stories had the best engagement and created positive comments [64]. These findings were also not surprising, as Heldman et al have suggested that soliciting user-generated content is something that should be encouraged to promote user engagement when using social media for public health purposes [39]. This suggests that use of personal stories and experiences is a valuable way to promote engagement in social media for health purposes.

Although featured story submissions had the highest engagement on Facebook and Instagram, using likes and comments, this was not the case with Twitter. On Twitter, the posters advertising the campaign had the highest engagement, and featured stories and supplemental content related to the featured story had similar engagement levels. These findings may be attributed to the 280-character limit imposed by Twitter. As a result of this character limit, the EWS team had to shorten and paraphrase the stories to fit in the allotted space. This paraphrasing often removed the bulk of the story and, as a result, may have diminished the personal connection that followers found in the featured story submissions on the other platforms. Another possible reason for this finding is that Facebook, Instagram, and Twitter are unique platforms, and each platform has a different user demographic [41]. For example, Instagram has a large number of users aged 18 to 24 years [41]. In addition, in a study examining preferences for a diabetes health promotion campaign, Gabarron et al [65] determined that Facebook was the most preferred platform overall for health promotion campaigns, and Twitter was the least preferred. These preferences could also explain different findings in the relative performance of different types of posts on the different platforms. It is important when developing and implementing social media campaigns for health-related purposes that the developers investigate and determine which platform is most appropriate to choose, as preferences change frequently and new platforms are always emerging [66]. Pilot studies and survey research of target followers may be a strategy to gather this information.

Overall, followers appreciated the positive messages and content of this campaign, especially in a time of increased exposure to negative news. This finding aligns with results found by Pahayahay and Khalili-Mahani [42] described earlier. Additionally, there is a growing concern surrounding the quality of information shared on social media, including the presence of nonevidence-based health information; therefore, recommendations have been made for health professionals to counteract this misinformation by providing simple, evidence-based content through social media channels [67]. Interview participants felt that the #eatwellcovid19 campaign shared valuable stories and reliable and attainable nutrition information and advice. It was reassuring for followers that content was evidence based, positive, and vetted by a dietitian. Different kinds of followers followed the campaign, so a variety of content is very important to appeal to everyone. Although some followers liked how the campaign helped to connect individuals when they were apart from others, others still followed the campaign for simple food and nutrition information, and not just for community connection. This should be taken into consideration when developing future health promotion campaigns.

EWS is a relatively new service in Saskatchewan, established in March 2019, and has since been actively growing its online social media following. As EWS's following continues to grow, the performance of future campaigns will likely improve, especially considering that 158,162 impressions were made on a combined total of only 2306 followers across the three social media platforms. Focusing on the social media metrics of the campaign, Facebook performed the best, which is not surprising because it is also EWS's most followed platform. The campaign made more impressions on Twitter than it did on Instagram, which is surprising because EWS has more followers on Instagram than Twitter and 19 more #eatwellcovid19 posts were shared on Instagram compared to Twitter. All platforms experienced a growth in their following, with 30% growth in EWS Instagram's following being the most substantial. The growth in followers across platforms because of the #eatwellcovid19 campaign suggests that health professionals and health organizations with a small social media following could use a health promotion social media campaign to establish themselves on social media and increase their following [62].

The design of our campaign could be viewed as a strength or a limitation, as the success of the campaign was dependent on the participation of EWS social media followers and their willingness to submit stories. Due to the high volume and quality of the submissions received, the #eatwellcovid19 campaign was

XSL•FO
RenderX

able to share numerous stories on a variety of topics as well as credible nutrition information that supplemented those stories. However, if EWS had only received a small number of stories for the campaign, there would not have been enough material to make it successful. Although personal stories are empowering and are a valuable tool for public engagement, it can be time-consuming to collect stories and turn them into posts, and it is possible that the content submitted by followers may not align well with the overall goals or message of the campaign. Therefore, health professionals who have specific content to share in a timely manner may have a difficult time incorporating user-shared stories into their campaign design.

A limitation of the #eatwellcovid19 evaluation was that we were unable to determine if eating behavior changes occurred as a direct result of the campaign; however, some qualitative data suggest that small behavior changes may have occurred or been prompted by the campaign. Another limitation of the study was that the interviewer for the semistructured qualitative interviews was not external to the campaign. In addition, we are unsure of the demographics of EWS followers, as social media metrics from each platform do not provide detailed follower demographic information (ie, where a follower lives). Therefore, we are also unsure if the participants who completed interviews are directly representative of campaign followers. A final limitation and a lesson learned for future social media campaigns is that detailed 24-hour story metrics are only available for a short time on Facebook and Instagram (eg, 28 days on Facebook). We only learned after the campaign was finished that Instagram and Facebook 24-hour story metrics could not be collected, as we collected our data past the time point of where this information was available. As a result, we were only able to capture the number of stories shared. This oversight was unfortunate because there was a substantial amount of reach and follower engagement that occurred with using the 24-hour stories in the #eatwellcovid19 campaign. In future campaigns, it would be valuable to record the number of stories shared and continually collect these metrics, such as reach and engagement, to better evaluate campaign performance. In addition, use of the 24-hour story feature in social media is a promising area of future research for health-related social media campaigns.

Conclusions

Numerous stories were shared to #eatwellcovid19 that showcased various strategies on how individuals were eating well during the COVID-19 pandemic. On Instagram and Facebook, posts that featured these stories had the highest engagement compared to other post types (eg, supplemental content related to the feature story). During this campaign, EWS's social media following increased by more than 10% on each platform. The #eatwellcovid19 campaign can be used as a reference for other health care professionals interested in the design of future social media campaigns.

Acknowledgments

The #eatwellcovid19 campaign was funded by the Saskatchewan Health Research Foundation Research Connections COVID-19 Rapid Response grant. We would like to thank all participants who completed the interviews to evaluate the campaign.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of social media content used in #eatwellcovid19.
[PDF File (Adobe PDF File), 504 KB - jmir_v23i7e27448_app1.pdf]

Multimedia Appendix 2
Interview protocol.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i7e27448_app2.pdf]

References

1. Listings of WHO's response to COVID-19. World Health Organization. 2020 Jun 29. URL: https://www.who.int/news/item/29-06-2020-covidtimeline [accessed 2020-07-15]
2. Coronavirus disease (COVID-19). World Health Organization. 2020 Oct 12. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19 [accessed 2020-07-15]
3. Modes of transmission of virus causing COVID-19: Implications for IPC precaution recommendations. World Health Organization. 2020 Mar 29. URL: https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations [accessed 2021-01-01]
4. Coronavirus disease (COVID-19): Prevention and risks. Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks.html [accessed 2021-01-01]
5. Individual and community-based measures to mitigate the spread of COVID-19 in Canada. Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/public-health-measures-mitigate-covid-19.html [accessed 2021-05-22]

XSL•FO
RenderX

6. Shahab S. Public Health Order (Under Subsection 45(2) of The Public Health Act, 1994 and section 25.2(2) of The Disease Control Regulations) Control of Transmission of 2019 Novel Coronavirus. Regina, SK: Government of Saskatchewan; 2021 Apr 07. URL: https://publications.saskatchewan.ca/#/products/112503 [accessed 2020-12-31]
7. Shahab S. Public Health Order (Under Subsection 38 and Subsection 45(2) of The Public Health Act, 1994) Control of Transmission of 2019 Novel Coronavirus. Regina, SK: Government of Saskatchewan; 2020 Mar 26. URL: https://publications.saskatchewan.ca/#/products/104652 [accessed 2020-12-30]
8. Schmidt C, Goetz S, Rocker S, Tian Z. Google searches reveal changing consumer food sourcing in the COVID-19 pandemic. J Agric Food Syst Community Dev 2020 May 21;9(3):9-16 [FREE Full text] [doi: 10.5304/jafscd.2020.093.032]
9. Canadian consumers prepare for COVID-19. Statistics Canada. 2020 Apr 08. URL: https://www150.statcan.gc.ca/n1/pub/62f0014m/62f0014m2020004-eng.htm [accessed 2021-05-28]
10. Martin-Neuninger R, Ruby M. What does food retail research tell us about the implications of coronavirus (COVID-19) for grocery purchasing habits? Front Psychol 2020;11:1448 [FREE Full text] [doi: 10.3389/fpsyg.2020.01448] [Medline: 32581987]
11. Deaton BJ, Deaton BJ. Food security and Canada's agricultural system challenged by COVID‐19. Can J Agric Econ 2020 Apr 26;68(2):143-149. [doi: 10.1111/cjag.12227]
12. Wolfson JA, Leung CW. Food insecurity and COVID-19: Disparities in early effects for US adults. Nutrients 2020 Jun 02;12(6):1648 [FREE Full text] [doi: 10.3390/nu12061648] [Medline: 32498323]
13. Ben Hassen T, El Bilali H, Allahyari MS, Berjan S, Fotina O. Food purchase and eating behavior during the COVID-19 pandemic: A cross-sectional survey of Russian adults. Appetite 2021 Oct 01;165:105309 [FREE Full text] [doi: 10.1016/j.appet.2021.105309] [Medline: 34015307]
14. Sarney G. CFDR study reveals consumer confusion related to nutrition information. Canadian Foundation for Dietetic Research. 2016 Jun 09. URL: https://www.cfdr.ca/Downloads/News-Releases/2016/Tracking-Nutrition-Trends-Media-Release--2016-June.aspx [accessed 2021-05-29]
15. Tracking nutrition trends 2015. Canadian Foundation for Dietetic Research. 2015. URL: https://www.cfdr.ca/Downloads/Tracking-Nutrition-Trends/Tracking-Nutrition-Trends-IX-2015.aspx [accessed 2021-05-29]
16. Ali SH, Imbruce VM, Russo RG, Kaplan S, Stevenson K, Mezzacca TA, et al. Evaluating closures of fresh fruit and vegetable vendors during the COVID-19 pandemic: Methodology and preliminary results using omnidirectional street view imagery. JMIR Form Res 2021 Feb 18;5(2):e23870 [FREE Full text] [doi: 10.2196/23870] [Medline: 33539310]
17. Richards TJ, Rickard B. COVID‐19 impact on fruit and vegetable markets. Can J Agric Econ 2020 May 18;68(2):189-194. [doi: 10.1111/cjag.12231]
18. Anelich LECM, Lues R, Farber JM, Parreira VR. SARS-CoV-2 and risk to food safety. Front Nutr 2020;7:580551 [FREE Full text] [doi: 10.3389/fnut.2020.580551] [Medline: 33224968]
19. Han J, Zhang X, He S, Jia P. Can the coronavirus disease be transmitted from food? A review of evidence, risks, policies and knowledge gaps. Environ Chem Lett 2021;19:5-16 [FREE Full text] [doi: 10.1007/s10311-020-01101-x] [Medline: 33024427]
20. COVID-19 and food safety: Guidance for food businesses. Interim guidance. World Health Organization. 2020 Apr 07. URL: https://www.who.int/publications/i/item/covid-19-and-food-safety-guidance-for-food-businesses [accessed 2021-05-29]
21. Kumar S, Singh R, Kumari N, Karmakar S, Behera M, Siddiqui AJ, et al. Current understanding of the influence of environmental factors on SARS-CoV-2 transmission, persistence, and infectivity. Environ Sci Pollut Res Int 2021 Feb;28(6):6267-6288 [FREE Full text] [doi: 10.1007/s11356-020-12165-1] [Medline: 33387315]
22. Adams EL, Caccavale LJ, Smith D, Bean MK. Food insecurity, the home food environment, and parent feeding practices in the era of COVID-19. Obesity (Silver Spring) 2020 Nov;28(11):2056-2063 [FREE Full text] [doi: 10.1002/oby.22996] [Medline: 32762129]
23. Men F, Tarasuk V. Food insecurity amid the COVID-19 pandemic: Food charity, government assistance, and employment. Can Public Policy 2021 Jun 01;47(2):202-230. [doi: 10.3138/cpp.2021-001]
24. Dondi A, Candela E, Morigi F, Lenzi J, Pierantoni L, Lanari M. Parents' perception of food insecurity and of its effects on their children in Italy six months after the COVID-19 pandemic outbreak. Nutrients 2020 Dec 31;13(1):121 [FREE Full text] [doi: 10.3390/nu13010121] [Medline: 33396310]
25. Niles MT, Bertmann F, Belarmino EH, Wentworth T, Biehl E, Neff R. The early food insecurity impacts of COVID-19. Nutrients 2020 Jul 15;12(7):2096 [FREE Full text] [doi: 10.3390/nu12072096] [Medline: 32679788]
26. Carroll N, Sadowski A, Laila A, Hruska V, Nixon M, Ma DW, on behalf of the Guelph Family Health Study. The impact of COVID-19 on health behavior, stress, financial and food security among middle to high income Canadian families with young children. Nutrients 2020 Aug 07;12(8):2352 [FREE Full text] [doi: 10.3390/nu12082352] [Medline: 32784530]
27. Noyes I, Lyle N. COVID-19 and school food: The impact of the early stages of the coronavirus pandemic on student nutrition programs in Ontario. J Agric Food Syst Community Dev 2021 Apr 06;10(2):197-210. [doi: 10.5304/jafscd.2021.102.049]
28. Lamarche B, Brassard D, Lapointe A, Laramée C, Kearney M, Côté M, et al. Changes in diet quality and food security among adults during the COVID-19-related early lockdown: Results from NutriQuébec. Am J Clin Nutr 2021 Apr 06;113(4):984-992 [FREE Full text] [doi: 10.1093/ajcn/nqaa363] [Medline: 33398347]

XSL•FO
RenderX

29. Bertrand L, Shaw KA, Ko J, Deprez D, Chilibeck PD, Zello GA. The impact of the coronavirus disease 2019 (COVID-19) pandemic on university students' dietary intake, physical activity, and sedentary behaviour. Appl Physiol Nutr Metab 2021 Mar;46(3):265-272 [FREE Full text] [doi: 10.1139/apnm-2020-0990] [Medline: 33449864]
30. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: Cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
31. van der Linden S, Roozenbeek J, Compton J. Inoculating against fake news about COVID-19. Front Psychol 2020;11:566790 [FREE Full text] [doi: 10.3389/fpsyg.2020.566790] [Medline: 33192844]
32. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020 Apr 01;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]
33. Adams KK, Baker WL, Sobieraj DM. Myth busters: Dietary supplements and COVID-19. Ann Pharmacother 2020 Aug;54(8):820-826. [doi: 10.1177/1060028020928052] [Medline: 32396382]
34. Health product advertising incidents related to COVID-19. Government of Canada. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/covid19-industry/health-product-advertising-incidents.html [accessed 2021-01-01]
35. Gleeson M, Nieman DC, Pedersen BK. Exercise, nutrition and immune function. J Sports Sci 2004 Jan;22(1):115-125. [doi: 10.1080/02640410310001405590] [Medline: 14971437]
36. Henrina J, Lim MA, Pranata R. COVID-19 and misinformation: How an infodemic fuelled the prominence of vitamin D. Br J Nutr 2021 Feb 14;125(3):359-360 [FREE Full text] [doi: 10.1017/S0007114520002950] [Medline: 32713358]
37. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
38. Dumas A, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: Scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55 [FREE Full text] [doi: 10.2196/jmir.9230] [Medline: 29463487]
39. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: Implications for public health organizations being truly “social”. Public Health Rev 2013 Jun 3;35(1):1-18 [FREE Full text] [doi: 10.1007/BF03391698]
40. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
41. Gruzd A, Mai P. The state of social media in Canada 2020. SSRN 2020 Jul 13:1-21 [FREE Full text] [doi: 10.2139/ssrn.3651206]
42. Pahayahay A, Khalili-Mahani N. What media helps, what media hurts: A mixed methods survey study of coping with COVID-19 using the media repertoire framework and the appraisal theory of stress. J Med Internet Res 2020 Aug 06;22(8):e20186 [FREE Full text] [doi: 10.2196/20186] [Medline: 32701459]
43. Graffigna G, Bosio C, Savarese M, Barello M, Barello S. "#I-Am-Engaged": Conceptualization and first implementation of a multi-actor participatory, co-designed social media campaign to raise Italians citizens' engagement in preventing the spread of COVID-19 virus. Front Psychol 2020;11:567101 [FREE Full text] [doi: 10.3389/fpsyg.2020.567101] [Medline: 33250811]
44. Lucini D, Gandolfi C, Antonucci C, Cavagna A, Valzano E, Botta E, et al. #StayHomeStayFit: UNIMI's approach to online healthy lifestyle promotion during the COVID-19 pandemic. Acta Biomed 2020 Sep 07;91(3):e2020037 [FREE Full text] [doi: 10.23750/abm.v91i3.10375] [Medline: 32921731]
45. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infect Control Hosp Epidemiol 2021 Feb;42(2):131-138 [FREE Full text] [doi: 10.1017/ice.2020.406] [Medline: 32758315]
46. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: A scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
47. Eat Well Saskatchewan. URL: https://eatwellsask.usask.ca/ [accessed 2021-06-29]
48. COVID-19 FAQ. Eat Well Saskatchewan. URL: https://eatwellsask.usask.ca/covid-19-faq.php [accessed 2021-06-29]
49. Census Profile, 2016 Census: Saskatchewan [Province] and Canada [Country]. Statistics Canada. URL: https://tinyurl.com/4whv5xuh [accessed 2020-07-15]
50. Saskatchewan Population Report: 2016 Census of Canada. Regina, SK: Saskatchewan Bureau of Statistics; 2017 Feb 08. URL: https://pubsaskdev.blob.core.windows.net/pubsask-prod/100723/100723-Saskatchewan_Population_Report_for_the_2016_Census.pdf [accessed 2020-10-15]
51. Saskatchewan Quarterly Population Report: Second Quarter 2020. Regina, SK: Saskatchewan Bureau of Statistics; 2020 Sep 29. URL: https://tinyurl.com/3uvf3bsx [accessed 2020-10-15]
52. What's the difference between Page views, reach and impressions on Facebook? Facebook Help Center. URL: https://www.facebook.com/help/274400362581037/?helpref=uf_permalink&parent_cms_id=794890670645072 [accessed 2021-05-29]

XSL•FO
RenderX

53. Post engagement. Facebook for Business: Business Help Center. URL: https://www.facebook.com/business/help/735720159834389 [accessed 2021-05-29]
54. Instagram post reach. Facebook for Business: Business Help Center. URL: https://www.facebook.com/business/help/430871940920146?helpref=search&sr=1&query=instagram%20reach [accessed 2021-05-29]
55. Tweet and video activity dashboards. Twitter Ads Help Center. URL: https://business.twitter.com/en/help/campaign-measurement-and-analytics/tweet-activity-dashboard.html [accessed 2021-05-29]
56. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
57. Patton MQ. Qualitative Research and Evaluation Methods. 3rd edition. Thousand Oaks, CA: SAGE Publications; 2002.
58. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
59. Creswell JW. Glossary. In: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th edition. Thousand Oaks, CA: SAGE Publications; 2014:248.
60. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
61. Graham JE, Moore JL, Bell RC, Miller T. Digital marketing to promote healthy weight gain among pregnant women in Alberta: An implementation study. J Med Internet Res 2019 Feb 01;21(2):e11534 [FREE Full text] [doi: 10.2196/11534] [Medline: 30707100]
62. Perrault EK, Hildenbrand GM, McCullock SP, Schmitz KJ, Dolick KN. Hashtag health: College health on social media and students' motivations to follow, interact, and share their social media content. Health Promot Pract 2019 Sep;20(5):721-729. [doi: 10.1177/1524839919853820] [Medline: 31169032]
63. Gabarron E, Larbi D, Dorronzoro E, Hasvold P, Wynn R, Årsand E. Factors engaging users of diabetes social media channels on Facebook, Twitter, and Instagram: Observational study. J Med Internet Res 2020 Sep 29;22(9):e21204 [FREE Full text] [doi: 10.2196/21204] [Medline: 32990632]
64. Pedersen EA, Loft LH, Jacobsen SU, Søborg B, Bigaard J. Strategic health communication on social media: Insights from a Danish social media campaign to address HPV vaccination hesitancy. Vaccine 2020 Jun 26;38(31):4909-4915. [doi: 10.1016/j.vaccine.2020.05.061] [Medline: 32482460]
65. Gabarron E, Dorronzoro E, Bradway M, Rivera-Romero O, Wynn R, Årsand E. Preferences and interests of diabetes social media users regarding a health-promotion intervention. Patient Prefer Adherence 2018;12:2499-2506 [FREE Full text] [doi: 10.2147/PPA.S184369] [Medline: 30538433]
66. Ortiz-Ospina E. The rise of social media. Our World in Data. 2019 Sep 18. URL: https://ourworldindata.org/rise-of-social-media [accessed 2021-05-29]
67. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
EWS: Eat Well Saskatchewan
FAQ: frequently asked questions

Edited by C Basch; submitted 29.01.21; peer-reviewed by L Mainvil, K Reuter; comments to author 09.03.21; revised version received 30.05.21; accepted 01.06.21; published 21.07.21.

Please cite as:
Grantham JL, Verishagen CL, Whiting SJ, Henry CJ, Lieffers JRL
Evaluation of a Social Media Campaign in Saskatchewan to Promote Healthy Eating During the COVID-19 Pandemic: Social Media Analysis and Qualitative Interview Study
J Med Internet Res 2021;23(7):e27448
URL: https://www.jmir.org/2021/7/e27448
doi:10.2196/27448
PMID:34133314

©Jordyn L Grantham, Carrie L Verishagen, Susan J Whiting, Carol J Henry, Jessica R L Lieffers. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.07.2021. This is an open-access article distributed under the

XSL•FO
RenderX

terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Risk Factors of Psychological Responses of Chinese University Students During the COVID-19 Outbreak: Cross-sectional Web-Based Survey Study

Xudong Zhang[1], MD; Xin Shi[2], PhD; Yang Wang[3], PhD; Huiquan Jing[4], PhD; Qingqing Zhai[5], PhD; Kunhang Li[1], MD; Dan Zhao[6], PhD; Shiyu Zhong[1], MD; Yuequn Song[1], MD; Feng Zhang[1], MD; Yijun Bao[1], MD, PhD

[1]Department of Neurosurgery, The Fourth Hospital of China Medical University, Shenyang, China

[2]Business School, All Saints Campus, Manchester Metropolitan University, Manchester, United Kingdom

[3]Department of Social Medicine, School of Public Health, China Medical University, Shenyang, China

[4]School of Public Health, Capital Medical University, Beijing, China

[5]School of Management, Shanghai University, Shanghai, China

[6]Department of Neurosurgery, The First Hospital of China Medical University, Shenyang, China

Corresponding Author:
Yijun Bao, MD, PhD
Department of Neurosurgery
The Fourth Hospital of China Medical University
No 4 Chongshandong
Huanggu
Shenyang, 110084
China
Phone: 86 15840150493
Email: yjbao@cmu.edu.cn

Abstract

Background: COVID-19 is a highly contagious and highly pathogenic disease caused by a novel coronavirus, SARS-CoV-2, and it has become a pandemic. As a vulnerable population, university students are at high risk during the epidemic, as they have high mobility and often overlook the severity of the disease because they receive incomplete information about the epidemic. In addition to the risk of death from infection, the epidemic has placed substantial psychological pressure on the public. In this respect, university students are more prone to psychological problems induced by the epidemic compared to the general population because for most students, university life is their first time outside the structure of the family, and their mental development is still immature. Internal and external expectations and academic stress lead to excessive pressure on students, and unhealthy lifestyles also deteriorate their mental health. The outbreak of COVID-19 was a significant social event, and it could potentially have a great impact on the life and the mental health of university students. Therefore, it is of importance to investigate university students' mental health status during the outbreak of COVID-19.

Objective: The principal objective of this study was to investigate the influencing factors of the psychological responses of Chinese university students during the COVID-19 outbreak.

Methods: This study used data from a survey conducted in China between February 21 and 24, 2020, and the data set contains demographic information and psychological measures including the Self-Rating Anxiety Scale, the Self-Rating Depression Scale, and the compulsive behaviors portion of the Yale-Brown Obsessive-Compulsive Scale. A total of 2284 questionnaires were returned, and 2270 of them were valid and were used for analysis. The Mann-Whitney *U* test for two independent samples and binary logistic regression models were used for statistical analysis.

Results: Our study surveyed 563 medical students and 1707 nonmedical students. Among them, 251/2270 students (11.06%) had mental health issues. The results showed that contact history of similar infectious disease (odds ratio [OR] 3.363, *P*=.02), past medical history (OR 3.282, *P*<.001), and compulsive behaviors (OR 3.525, *P*<.001) contributed to the risk of mental health issues. Older students (OR 0.928, *P*=.02), regular daily life during the epidemic outbreak (OR 0.410, *P*<.001), exercise during the epidemic outbreak (OR 0.456, *P*<.001), and concern related to COVID-19 (OR 0.638, *P*=.002) were protective factors for mental health issues.

XSL•FO
RenderX

Conclusions: According to the study results, mental health issues have seriously affected university students, and our results are beneficial for identifying groups of university students who are at risk for possible mental health issues so that universities and families can prevent or intervene in the development of potential mental health issues at the early stage of their development.

(J Med Internet Res 2021;23(7):e29312) doi:10.2196/29312

KEYWORDS

university students; depressive symptoms; anxiety symptoms; mental health status; COVID-19; pandemic; mental health; anxiety; psychological health

Introduction

A novel coronavirus pneumonia disease, COVID-19, spread very quickly across China in early 2020 [1]. The outbreak was first discovered in late December 2019, when a series of unexplained pneumonia cases were identified that were related to epidemiologically undiscovered seafood market exposure in Wuhan City, Hubei Province, China [2]. According to the official website of the National Health Commission of the People's Republic of China, on July 31, 2020, a total of 84,337 confirmed cases had accumulated, including 78,989 discharged cases and 4634 deaths; a total of 789,742 close contacts were tracked, and 20,278 close contacts were still under medical observation [3]. The World Health Organization declared the COVID-19 outbreak a public health emergency of international concern on January 30, 2020 [4]. More than 83 million confirmed infection cases across more than 200 countries globally had been reported as of December 31, 2020, including over 1.82 million deaths [5]. The advanced and rapid spread of COVID-19 has brought many complex challenges to the global public health and medical communities.

COVID-19 has brought the risk of death from infection and unbearable psychological pressure to people worldwide [6]. Previous research has shown that the COVID-19 has had a broad psychosocial impact on humans at the individual level, where people may feel fear of illness or death, helplessness, and stigma [7]. During a public health emergency, approximately 10% to 30% of the public are very concerned or quite concerned about the possibility of contracting the disease [8]. As schools and businesses are closed due to the pandemic, individuals' negative emotions become more complicated [9]. Many studies investigated the psychological impact on uninfected communities during the severe acute respiratory syndrome (SARS) outbreak and found a significant mental illness incidence [10]. Studies also found some risk factors for the deterioration of mental health: being female [11,12], having a medical history, such as prior psychiatric illness, physical illness, or chronic disease [13-15], being unmarried [16-18], having lower income [13,16,19], and experiencing a negative economic impact [20]. People who were more likely to take preventive measures against the infection were older, female, and highly educated; they also had intensive awareness of SARS, a moderate level of anxiety, and a positive contact history [21].

An essential part of the population is university students, who are under heavy study pressure and have unhealthy lifestyles; more importantly, the university stage is a stage of transition to maturity in life development. During university studies, the social knowledge that students have acquired may be insufficient to understand the pandemic due to the limited social activities of students compared with that of working adults [22]. Therefore, the students are not inclined to find ways to release pressure, which can lead to unstable mental states, and the situation can worsen in epidemics such as the COVID-19 pandemic. Besides, the sources of COVID-19 infection may be study places at the university, and populations of hundreds of millions of students are at risk of spreading the virus [23]. Many studies have shown that the outbreak of infectious diseases will have a psychological impact on the general population, including medical staff and university students. A prominent example is the mental sequelae observed during the outbreak of SARS in 2003 [24]. Studies of the SARS outbreak have shown that medical staff experienced acute stress reactions [25,26]. However, medical students also require attention because they are students with fragile mental endurance and medical workers without complete medical training, and their exposure risks are higher than those of other people.

Therefore, the aim of our study is to investigate the mental health of university students and its influencing factors during the COVID-19 pandemic on groups of both medical and nonmedical students. In this study, we conducted a survey of the targeted sample with the aims to explore risk factors that contributed to mental disease during the pandemic and to provide evidence for psychological intervention programs for university students. This research is essential for students' healthy growth and is an effective response to future work and mental health interventions for students.

Methods

Study Population and Sample

The targeted population included medical students from 57 universities in China. In this study, a cross-sectional survey was developed, and anonymous web-based questionnaires were used to investigate students' mental health status during the COVID-19 epidemic. A snowball sampling strategy was used; the web-based survey was first distributed to medical students, and they were encouraged to pass it on to others. A total of 2284 questionnaires were returned, and 14 of them were excluded because the respondents did not fill in the answers completely or did not meet the criteria of the survey; for example, some respondents were teachers and not students. A total of 2270 valid questionnaires were finalized in the study, including surveys from 563 medical students and 1707 nonmedical students.

XSL•FO
RenderX

Study Instruments

The questionnaire contained the demographic information and psychological measures, which include the Self-Rating Anxiety Scale (SAS), the Self-Rating Depression Scale (SDS), and the compulsive behaviors part of the Yale-Brown Obsessive-Compulsive Scale (YBOCS).

Demographic Information

The questions on demographic information in this study were related to gender, age, whether the respondent is an only child, ethnicity, place of residence, region of residence, whether the respondent has participated in volunteer work, contact history of similar infectious disease, past medical history, regularity of daily life and exercise during the epidemic, and concern about COVID-19. The geographical distribution map of the participants was depicted by ArcGIS software (Esri), and the cutoff points for classification were based on the Jenks classification technique as employed in ArcGIS, version 10.5.

The SAS and SDS

The SAS and SDS [27,28] were developed by William WK Zung, a psychiatrist at Duke University. Both the SAS and SDS are Likert scale surveys. They each contain 20 items of self-report examination that measure the level of anxiety symptoms (SAS) or depressive symptoms (SDS). The scores of the 20 items are added in each scale and then converted into standard scores, where higher scores represent more severe anxiety or depression. Based on the Chinese norm, people who score more than 50 on the SAS scale are defined as having anxiety symptoms, and people who score more than 53 on the SDS are treated as having depressive issues. The SAS demonstrated good internal consistency (Cronbach α=.828), and the SDS also showed good internal consistency (Cronbach α=.849).

The YBOCS

The YBOCS was developed to remedy problems with existing rating scales by providing a specific measure of the severity of symptoms of obsessive-compulsive disorder (OCD) [29]. This measure ensures that the mental disorder will not be influenced by the type of obsessions or compulsions present. The scale is clinician rated and includes 10 items; a score >6 indicates compulsive behavior. In this study, we used only part of the compulsive behaviors scale of the YBOCS, and it demonstrated good internal consistency (Cronbach α=.810).

Data Analysis

Data were analyzed using SPSS, version 26.0 (IBM Corporation). According to the data type, the Mann-Whitney *U* test, a type of nonparametric test, was used to explore the significant associations between sample characteristics and mental health issues during the COVID-19 epidemic. Binary logistic regression analysis was conducted for dependent variables (mental health issues) and independent variables (demographic and psychological measures), and we set the level of statistical significance as $P<.05$.

Ethics Approval and Consent to Participate

This study was conducted in compliance with the Declaration of Helsinki's ethical principles and its later amendments, and the study was reviewed and approved by the Ethics Committee on Human Experimentation of China Medical University (EC-2020-KS-025). The study procedures followed ethical standards. The participants were informed of the study protocol, and consent was received from all the participants. All participation was voluntary and anonymous. Confidentiality was ensured in processing personal data and maintaining individual records.

Results

The characteristics and other demographic information in the valid sample of 2270 respondents are summarized in Table 1. In this study, a new dependent variable (mental health issues) was created to evaluate which independent variables affected the mental health of university students, and the samples with positive mental health issues included students with both positive anxiety symptoms and positive depression symptoms. The results showed that 251 out of 2270 students (11.1%) had mental health issues; among these 251 students, 106 were male (42.2%) and 145 were female (57.8%). Of these 251 students, 214 (85.3%) were between 19 and 24 years of age. The distribution of participants covered the country of China, as shown in Figure 1.

The Mann-Whitney *U* test shows that age, contact history of similar infectious disease, past medical history, compulsive behaviors, the regularity of daily life during the epidemic outbreak, exercise during the epidemic outbreak, and concern about COVID-19 were correlated with mental health issues (all $P<.05$), as shown in Table 2. However, gender, being an only child, ethnicity, place of residence, region, joining in volunteer work, and student type (medical vs nonmedical) had no statistically significant associations with mental health issues (all $P>.05$).

XSL•FO
RenderX

Table 1. Distribution of anxiety symptoms, depressive symptoms, and mental health issues among students (N=2270).

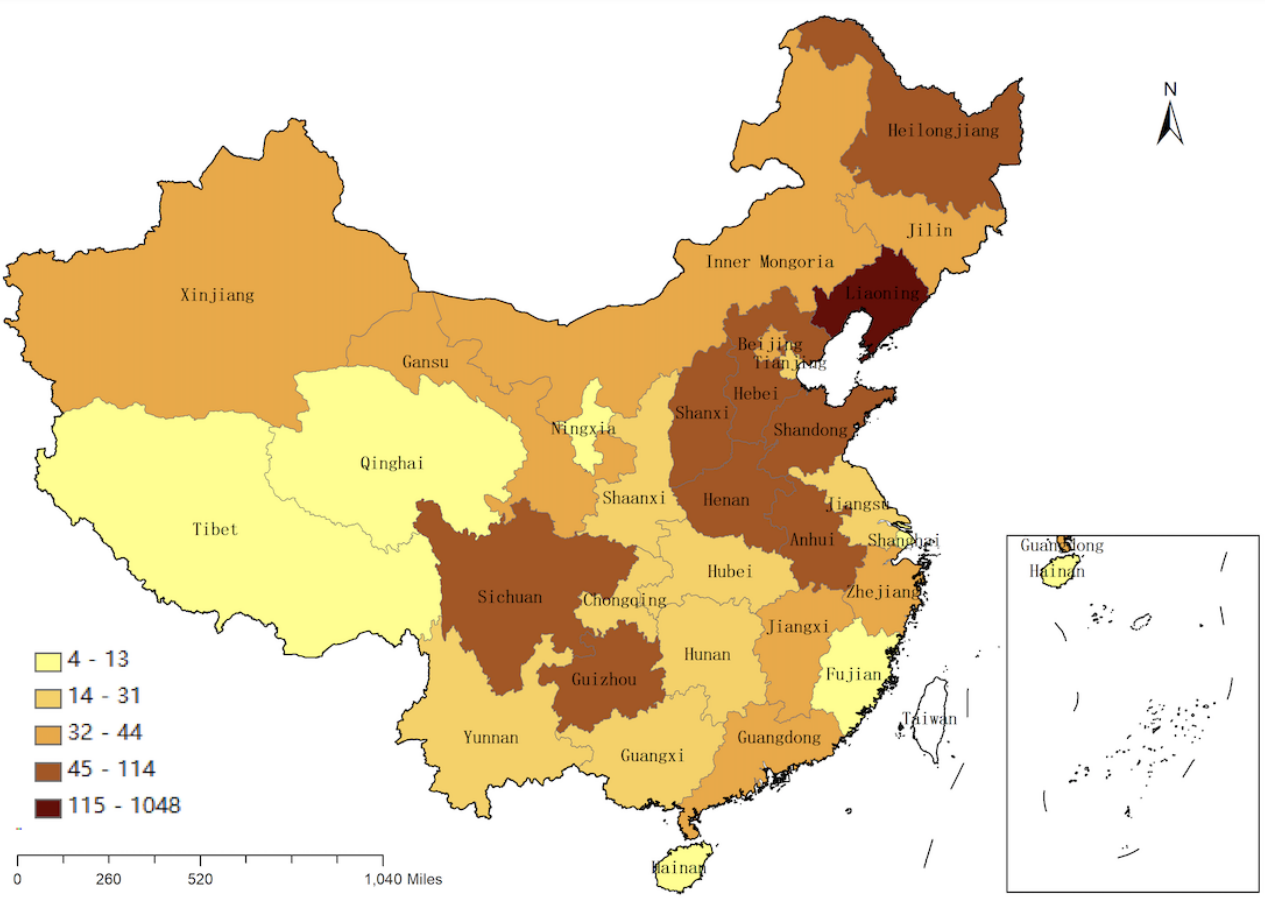
Variable	Total, n (%)[a]	Anxiety symptoms, n (%)		Depressive symptoms, n (%)		Mental health issues, n (%)	
		Positive	Negative	Positive	Negative	Positive	Negative
Gender							
Male	877 (38.6)	47 (5.4)	830 (94.6)	96 (10.9)	781 (89.1)	106 (12.1)	771 (87.9)
Female	1393 (61.4)	39 (2.8)	1354 (97.2)	141 (10.1)	1252 (89.9)	145 (10.4)	1248 (89.6)
Age (years)							
<18	250 (11)	10 (4)	240 (96)	31 (12.4)	219 (87.6)	33 (13.2)	217 (86.8)
19-24	1926 (84.8)	73 (3.8)	1853 (96.2)	204 (10.6)	1722 (89.4)	214 (11.1)	1712 (88.9)
≥25	94 (4.1)	3 (3.2)	91 (96.8)	2 (2.1)	92 (97.9)	4 (4.3)	90 (95.7)
Only child							
Yes	1051 (46.3)	43 (4.1)	1008 (95.9)	112 (10.7)	939 (89.3)	120 (11.4)	931 (88.6)
No	1219 (53.7)	43 (3.5)	1176 (96.5)	125 (10.3)	1094 (89.7)	131 (10.7)	1088 (89.3)
Ethnicity							
Han	1936 (85.3)	76 (3.9)	1860 (96.1)	206 (10.6)	1730 (89.4)	215 (11.1)	1721 (88.9)
Minority	334 (14.7)	10 (3.0)	324 (97.0)	31 (9.3)	303 (90.7)	36 (10.8)	298 (89.2)
Place of residence							
Urban	938 (41.3)	34 (3.6)	904 (96.4)	98 (10.4)	840 (89.6)	103 (11.0)	835 (89.0)
Rural	1332 (58.7)	52 (3.9)	1280 (96.1)	139 (10.4)	1193 (89.6)	148 (11.1)	1184 (88.9)
Region							
Hubei Province	26 (1.1)	1 (3.8)	25 (96.2)	2 (7.7)	24 (92.3)	2 (7.7)	24 (92.3)
Outside Hubei Province	2244 (99)	85 (3.8)	2159 (96.2)	235 (10.5)	2009 (89.5)	249 (11.1)	1995 (88.9)
Joined in volunteer work							
Yes	246 (10.8)	9 (3.7)	237 (96.3)	30 (12.2)	216 (87.8)	31 (12.6)	215 (87.4)
No	2024 (89.2)	77 (3.8)	1947 (96.2)	207 (10.2)	1817 (89.8)	220 (10.9)	1804 (89.1)
Contact history of similar infectious disease							
Yes	23 (1)	4 (17.4)	19 (82.6)	8 (34.8)	15 (65.2)	8 (34.8)	15 (65.2)
No	2247 (99)	82 (3.6)	2165 (96.4)	229 (10.2)	2018 (89.8)	243 (10.8)	2004 (89.2)
Past medical history							
Yes	101 (4.4)	12 (11.9)	89 (88.1)	27 (26.7)	74 (73.3)	29 (28.7)	72 (71.3)
No	2169 (95.6)	74 (3.4)	2095 (96.6)	210 (9.7)	1959 (90.3)	222 (10.2)	1947 (89.8)
Compulsive behaviors							
Yes	313 (13.8)	35 (11.2)	278 (88.8)	76 (24.3)	237 (75.7)	81 (25.9)	232 (74.1)
No	1957 (89.2)	51 (2.6)	1906 (97.4)	161 (8.2)	1796 (91.8)	170 (8.7)	1787 (91.3)
Regularity of daily life							
Regular	1301 (57.3)	19 (1.5)	1282 (98.5)	76 (5.8)	1225 (94.2)	79(6.1)	1222 (93.9)
Irregular	969 (42.7)	67 (6.9)	902 (93.1)	161 (16,6)	808 (83.4)	172 (17.8)	797 (82.2)
Exercise							
No exercise	809 (35.6)	47 (5.8)	762 (94.2)	135 (16.7)	674 (83.3)	143 (17.7)	666 (82.3)
Continued exercise	1461 (64.4)	39 (2.7)	1422 (97.3)	102 (7.0)	1359 (93.0)	108 (7.4)	1353 (92.6)
Concern about COVID-19							
Not very concerned (<1 hour per day)	1036 (45.6)	32 (3.1)	1004 (96.9)	131 (12.6)	905 (87.4)	135 (13.0)	901 (87.0)

XSL•FO
RenderX

Variable	Total, n (%)[a]	Anxiety symptoms, n (%)		Depressive symptoms, n (%)		Mental health issues, n (%)	
		Positive	Negative	Positive	Negative	Positive	Negative
Very concerned (>1 hour per day)	1234 (54.4)	54 (4.4)	1180 (95.6)	106 (8.6)	1128 (91.4)	116 (9.4)	1118 (90.6)
Student type							
Medical student	563 (24.8)	20 (3.6)	543 (96.4)	57 (10.1)	506 (89.9)	60 (10.7)	503 (89.3)
Nonmedical student	1707 (75.2)	66 (3.9)	1641 (96.1)	180 (10.5)	1527 (89.5)	191 (11.2)	1516 (88.8)
Total	2270 (100)	86 (3.8)	2184 (96.2)	237 (10.4)	2033 (89.6)	251 (11.1)	2019 (88.9)

[a]Percentages in the Total column are calculated based on N=2270; all other percentages are calculated based on the values in the Total column.

Figure 1. Geographical distribution map of the 2270 study participants.



XSL•FO
RenderX

Table 2. Mann-Whitney *U* tests and z scores of factors affecting students' mental health issues as the grouping variable.

Variable (factor)	Mann-Whitney *U*	Mann-Whitney *z*	*P* value
Gender (X8)	243138.00	–1.241	.22
Age (X1)	241098.00	–2.015	.04
Only child (X11)	249085.00	–0.508	.61
Ethnicity (X13)	252327.50	–0.176	.86
Place of residence (X14)	252570.50	–0.097	.92
Region (X10)	252391.50	–0.550	.58
Joining in volunteer work (X9)	249072.50	–0.818	.41
Contact history of similar infectious disease (X2)	247191.00	–3.646	<.001
Past medical history (X3)	233145.00	–5.787	<.001
Compulsive behaviors (X4)	200731.00	–9.003	<.001
Regularity of daily life (X5)	179774.00	–8.774	<.001
Exercise (X6)	192609.00	–7.481	<.001
Concern about COVID-19 (X7)	230177.50	–2.747	.006
Student type (X12)	250828.00	–0.349	.73

The above 7 factors (age, contact history of similar infectious disease, past medical history, compulsory behaviors, regularity of daily life, exercise, and concern about COVID-19) that had a significant association (P<.05) in the Mann-Whitney *U* test were entered into binary logistic regression model 1 as independent variables for mental health issues. Likewise, all the other above factors were entered into the binary logistic regression model in order of *P* value from small to large (the order is indicated in Table 2). First, we analyzed the overall effectiveness of these models (Table 3). The original hypothesis of the model test was that the quality of the model would be the same if the independent variable was or was not included. The results showed that the *P* value is <.05, which means that the original hypothesis is rejected, and the construction of these models is meaningful; moreover, based on the Akaike information criterion (AIC) goodness-of-fit statistic for comparing models, model 1 (AIC=1397.629), with the lowest AIC statistic, is the model with the best fit.

Table 3. Model summary of binary logistic regression analysis of students' mental health issues.

Model	–2 log likelihood	Chi-square (*df*)	*P* value	AIC[a]	BIC[b]
Intercept-only	1578.608	N/A[c]	N/A	N/A	N/A
Model 1 (X1, X2, X3, X4, X5, X6, X7)	1381.629	196.979 (7)	<.001	1397.629	1443.449
Model 2 (X1 ,X2, X3, X4, X5, X6, X7, X8)	1380.594	198.014 (8)	<.001	1398.594	1450.142
Model 3 (X1, X2, X3, X4, X5, X6, X7, X8, X9)	1378.846	199.762 (9)	<.001	1398.846	1456.121
Model 4 (X1, X2, X3, X4, X5, X6, X7, X8, X9, X10)	1378.483	200.125 (10)	<.001	1400.483	1463.486
Model 5 (X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11)	1378.313	200.295 (11)	<.001	1402.313	1471.044
Model 6 (X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12)	1377.973	200.635 (12)	<.001	1403.973	1478.431
Model 7 (X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13)	1377.778	200.830 (13)	<.001	1405.778	1485.963
Model 8 (X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14)	1377.540	201.068 (14)	<.001	1407.540	1493.453

[a]AIC: Akaike information criterion.

[b]BIC: Bayesian information criterion.

[c]N/A: not applicable.

As shown in Table 4, the above 7 factors in the optimal model (Model 1) were used as independent variables, and mental health issues were used as the dependent variable for binary logistic regression analysis; the formula of the model was $\square = 0.618 - 0.075 \times X1 + 1.213 \times X2 + 1.188 \times X3 + 1.260 \times X4 - 0.892 \times X5 - 0.786 \times X6 - 0.450 \times X7$ (where *P* represents the probability that mental health issues are positive and $1 - P$ represents the probability that mental health issues are negative). The analysis indicates that age (odds ratio [OR] 0.928, P=.02), regular daily life during the epidemic outbreak (OR=0.410, P<.001), exercise during the epidemic outbreak (OR 0.456,

XSL•FO
RenderX

P<.001), and concern about COVID-19 (OR 0.638, *P*=.002) were protective factors for mental health issues, as shown in Table 3. The results showed that older students and students who maintained regular daily life activity and physical exercise during the COVID-19 pandemic were less likely to have mental health issues. However, contact history of similar infectious disease (OR 3.363, *P*=.02), past medical history (OR 3.282, *P*<.001), and compulsive behaviors (OR 3.525, *P*<.001) were risk factors for mental health issues; this finding showed that students with these three conditions were more likely to have mental health issues.

Table 4. Estimations of binary logistic regression analysis of students' mental health issues as the dependent variable.

Variable	Mental health issues				
	B	SE	*z*	*P* value	Exp(B) (95% CI)
Age	–0.075	0.033	–2.301	.02	0.928 (0.870-0.989)
Contact history of similar infectious disease (yes vs no)	1.213	0.501	2.422	.02	3.363 (1.260-8.976)
Past medical history (yes vs no)	1.188	0.251	4.736	<.001	3.282 (2.007-5.367)
Compulsive behaviors (yes vs no)	1.260	0.167	7.554	<.001	3.525 (2.542-4.887)
Regularity of daily life (regular vs irregular)	–0.892	0.150	–5.936	<.001	0.410 (0.305-0.550)
Exercise (yes vs no)	–0.786	0.144	–5.446	<.001	0.456 (0.344-0.605)
Concern about COVID-19 (yes vs no)	–0.450	0.147	–3.066	.002	0.638 (0.478-0.850)
Constant	0.618	0.691	0.894	.37	1.855 (0.479-7.182)

Discussion

Principal Findings

In our study, we analyzed the psychological responses and associated factors, including risk factors and protective factors, of both medical and nonmedical students after the outbreak of COVID-19. We found that past medical history, contact history of similar infectious diseases, and compulsive behaviors contributed to the risk of mental health issues. Maintenance of regular daily life and exercise during the epidemic outbreak, older age, and high levels of concern (>1 hour per day) about COVID-19 were protective factors for mental health issues.

The COVID-19 outbreak is the largest outbreak of atypical pneumonia since SARS in 2003 in terms of the number of infection cases and time of spread [30]. COVID-19, compared with SARS, has brought greater risk of death and very high psychological pressure to people worldwide due to its power of "superspreading" among humans [6,31]. Studies have demonstrated the psychological impact of the early stage of COVID-19 on the general population, including college students [32-34], and it has been indicated that both medically trained medical staff and nonmedical health care personnel can be affected [35,36]. Our study provided evidence that 251/2270 students (11.1%) encountered mental health issues during the outbreak of COVID-19. During the continuous spread of the epidemic, strict isolation measures and closures in campuses across China may have affected university students' mental health [37,38]. More importantly, medical students who are medical workers with incomplete training possess only basic medical knowledge and do not have proficient professional skills and abundant clinical experience. Therefore, their clinical exposure risks are higher than those of other people. Likewise, the mental health of medical students requires attention during the pandemic.

A few studies reporting the psychological conditions of college students or medical students during the COVID-19 outbreak emerged during the preparation of our manuscript. Copeland et al [39] investigated the impact of the COVID-19 pandemic on the emotions, behaviors, and wellness behaviors of first-year college students and showed that COVID-19 and related mitigation strategies have a moderate but continuous impact on mood and healthy behavior. Bolatov et al [40] compared the mental state of medical students switching to web-based learning with that of students who received traditional learning during the COVID-19 pandemic, and they revealed that the prevalence of burnout syndrome, depression, anxiety, and somatic symptoms decreased after the transition. Li et al [41] investigated the rates of three mental health problems (acute stress, anxiety, and depressive symptoms) and their change patterns in two phases of the pandemic (early vs under control), and they showed that the significant predictors of distinct mental health trajectories included senior students, COVID-19 exposure, COVID-19–related worries, social support, and family function. Isralowitz et al [42] examined COVID-19–related fear and its association with psychoemotional conditions, including use of substances such as tobacco, alcohol, and cannabis, among Israeli and Russian social work students at two peak points or waves of infection. These literature reports focused on a specific population or condition, a relatively small sample size, and limited survey items during the COVID-19 pandemic. Compared with the above studies, our survey covered a large population of both medical and nonmedical university students, from undergraduates to graduates, with a wide geographical range using multidimensional survey items, including detailed demographic information and mental and behavioral status, and we analyzed the factors associated with the psychological responses. Our results are more specific and detailed.

Our study focused on the psychological status of university students and concluded that university students with compulsive behaviors are more likely to have mental health issues, which

XSL•FO
RenderX

confirmed the conclusions of previous research. The research has suggested that youth with OCD are at risk of experiencing comorbid psychiatric conditions, such as depression and anxiety [43]. Students with past medical history are more likely to have anxiety and depression symptoms; this is consistent with recent research findings, in which a medical history of issues such as prior psychiatric illness, physical illness, or chronic disease was a risk factor for the deterioration of mental health [13-15], and indicates that these students are more sensitive to the epidemic and require more psychological intervention.

University students with irregular daily life during the epidemic outbreak were more likely to have mental health issues, which is similar to previous research results. Previous studies have suggested that with the increasing pressure of modern life and irregular lifestyles, depression has become an increasing threat to human health. Studies have also suggested that better mental health at baseline was predicted by a lower body mass index, a higher frequency of physical and mental activities, nonsmoking, a nonvegetarian diet, and a more regular social rhythm [44]. Students should maintain a regular and healthy lifestyle during the epidemic to ensure good mental health status.

Our study also indicates that students with a contact history of similar infectious diseases are more likely to have mental health issues because this experience may cause students to worry about whether they have been infected, and medical students are more likely to be exposed to similar diseases in clinical work. Previous studies also suggested a significant association with anxiety for people whose contact histories included contact with an individual with suspected COVID-19 or with infected materials [31].

This study indicates that students who exercised during the epidemic outbreak were less likely to have mental health issues. Studies have suggested that physical exercise is associated with greater cardiovascular fitness, improved muscle strength and endurance, and reduction of depression and anxiety [45]. Students should continue to exercise during the epidemic to ensure good mental health. Unexpectedly, a high level of concern about COVID-19 was less likely to be associated with mental health issues owing to the dissemination of positive scientific information on Chinese media's public emergencies. Therefore, it is recommended to pay suitable attention to the news, especially the good news related to the epidemic, and such behaviors are beneficial to maintain a good attitude during the outbreak of COVID-19.

Previous studies have suggested that female and older students are more likely to have mental health symptoms [11,12]; women and older people have been found to experience more significant psychological impact and higher stress levels, anxiety, and depression [31]. These findings are inconsistent with our research results; because older students experienced SARS in 2003, they may have a more comprehensive understanding and a higher level of awareness of COVID-19 and may be less likely to have mental symptoms during the epidemic.

For family and society, these risk variables that cause mental health issues are key factors for early judgment of university students' psychological problems, and the results also provide the theoretical basis for formulating intervention measures. Schools, families, and the government should provide more care and support to university students during the epidemic.

Limitations

Given the limited available resources and the time of the COVID-19 outbreak, the study adopted the snowball sampling strategy, which is not based on randomly selected samples. Additionally, the researchers did not conduct a prospective study that would provide a specific measure to support the needs of targeted public health initiatives.

Conclusions

During the outbreak of COVID-19, some university students experienced mental symptoms. Past medical history, contact history of similar infectious disease, and compulsive behaviors were risk factors for mental health issues. Older age of the students, regular daily life, and exercise during the epidemic outbreak were protective factors against mental health issues. A high level of concern (>1 hour per day) about COVID-19 was also a protective factor.

These findings are beneficial for identifying the groups of university students at risk for possible mental health issues, and they provide a theoretical foundation for the formulation of relevant interventions so that universities and families can prevent or intervene in the development of mental health issues among students at the early stage of the disease. Likewise, the findings are essential for education and public health epidemic prevention. In short, students require more attention, help, and support from society, families, and universities during the COVID-19 pandemic.

Availability of Data and Materials

The data and materials used in this study are available upon request from the author.

Acknowledgments

This study was funded by the China Medical University novel coronavirus pneumonia prevention and control research project (2020-12-11); however, the views expressed are those of the authors and not necessarily those of the project. The study was also supported by the Liaoning Provincial Natural Science Foundation (2020-MS-155), Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (2013-1792), Fund for Scientific Research of The First Hospital of China Medical University (FSFH201722), and National Science Foundation of China (72074104). The researchers are grateful for the support of several organizations.

XSL•FO
RenderX

Authors' Contributions

YB, HQJ, and YW designed the web-based questionnaires. YB, XZ, KHL, DZ, SYZ, YQS, and FZ organized the data. XZ, XS, and QQZ analyzed and verified the data. XZ wrote the first draft. XS revised and polished the article. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
2. Nishiura H, Jung S, Linton NM, Kinoshita R, Yang Y, Hayashi K, et al. The extent of transmission of novel coronavirus in Wuhan, China, 2020. JCM 2020 Jan 24;9(2):330. [doi: 10.3390/jcm9020330]
3. The latest situation of the novel coronavirus pneumonia epidemic as of 24:00 on July 31. Webpage in Chinese. National Health Commission, People's Republic of China. 2020 Jul 31. URL: http://www.nhc.gov.cn/yjb/s7860/202008/96184996b6724df787ea9380dfdc88bf.shtml [accessed 2021-07-14]
4. Mahase E. China coronavirus: WHO declares international emergency as death toll exceeds 200. BMJ 2020 Jan 31:m408. [doi: 10.1136/bmj.m408]
5. Mallick S. Global Covid-19 cases top 83.9 million, deaths over 1.82 million. Odisha TV. 2021 Jan 02. URL: https://odishatv.in/coronavirus/global-covid-19-cases-top-83-9-million-deaths-over-1-82-million-505942 [accessed 2021-07-14]
6. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):300-302. [doi: 10.1016/s2215-0366(20)30073-0]
7. Hall R, Hall R, Chapman M. The 1995 Kikwit Ebola outbreak: lessons hospitals and physicians can apply to future viral epidemics. Gen Hosp Psychiatry 2008;30(5):446-452 [FREE Full text] [doi: 10.1016/j.genhosppsych.2008.05.003] [Medline: 18774428]
8. Rubin G, Potts H, Michie S. The impact of communications about swine flu (influenza A H1N1v) on public responses to the outbreak: results from 36 national telephone surveys in the UK. Health Technol Assess 2010 Jul;14(34):183-266 [FREE Full text] [doi: 10.3310/hta14340-03] [Medline: 20630124]
9. Van Bortel T, Basnayake A, Wurie F, Jambai M, Koroma AS, Muana AT, et al. Psychosocial effects of an Ebola outbreak at individual, community and international levels. Bull World Health Organ 2016 Mar 01;94(3):210-214 [FREE Full text] [doi: 10.2471/BLT.15.158543] [Medline: 26966332]
10. Sim K, Huak Chan Y, Chong PN, Chua HC, Wen Soon S. Psychosocial and coping responses within the community health care setting towards a national outbreak of an infectious disease. J Psychosom Res 2010 Feb;68(2):195-202 [FREE Full text] [doi: 10.1016/j.jpsychores.2009.04.004] [Medline: 20105703]
11. Lin LY, Wang J, Ou-Yang XY, Miao Q, Chen R, Liang FX, et al. The immediate impact of the 2019 novel coronavirus (COVID-19) outbreak on subjective sleep status. Sleep Med 2021 Jan;77(3):348-354 [FREE Full text] [doi: 10.1016/j.sleep.2020.05.018] [Medline: 32593614]
12. Gómez-Salgado J, Andrés-Villas M, Domínguez-Salas S, Díaz-Milanés D, Ruiz-Frutos C. Related health factors of psychological distress during the COVID-19 pandemic in Spain. Int J Environ Res Public Health 2020 Jun 02;17(11) [FREE Full text] [doi: 10.3390/ijerph17113947] [Medline: 32498401]
13. Ping W, Zheng J, Niu X, Guo C, Zhang J, Yang H, et al. Evaluation of health-related quality of life using EQ-5D in China during the COVID-19 pandemic. PLoS One 2020 Jun 18;15(6):e0234850 [FREE Full text] [doi: 10.1371/journal.pone.0234850] [Medline: 32555642]
14. Özdin S, Bayrak Özdin Ş. Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic in Turkish society: the importance of gender. Int J Soc Psychiatry 2020 Aug 08;66(5):504-511 [FREE Full text] [doi: 10.1177/0020764020927051] [Medline: 32380879]
15. Varshney M, Parel JT, Raizada N, Sarin SK. Initial psychological impact of COVID-19 and its correlates in Indian Community: An online (FEEL-COVID) survey. PLoS One 2020;15(5):e0233874 [FREE Full text] [doi: 10.1371/journal.pone.0233874] [Medline: 32470088]
16. Naser AY, Dahmash EZ, Al-Rousan R, Alwafi H, Alrawashdeh HM, Ghoul I, et al. Mental health status of the general population, healthcare professionals, and university students during 2019 coronavirus disease outbreak in Jordan: a cross-sectional study. Brain Behav 2020 Aug;10(8):e01730 [FREE Full text] [doi: 10.1002/brb3.1730] [Medline: 32578943]
17. Wang H, Xia Q, Xiong Z, Li Z, Xiang W, Yuan Y, et al. The psychological distress and coping styles in the early stages of the 2019 coronavirus disease (COVID-19) epidemic in the general mainland Chinese population: a web-based survey. PLoS One 2020;15(5):e0233410 [FREE Full text] [doi: 10.1371/journal.pone.0233410] [Medline: 32407409]
18. Fitzpatrick KM, Harris C, Drawve G. How bad is it? Suicidality in the middle of the COVID-19 pandemic. Suicide Life Threat Behav 2020 Dec;50(6):1241-1249 [FREE Full text] [doi: 10.1111/sltb.12655] [Medline: 32589799]

XSL•FO
RenderX

19. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-19 epidemic in southwestern China. Med Sci Monit 2020 Apr 20;26. [doi: 10.12659/msm.924609]
20. Guo J, Feng X, Wang X, van IJzendoorn MH. Coping with COVID-19: exposure to COVID-19 and negative impact on livelihood predict elevated mental health problems in Chinese adults. Int J Environ Res Public Health 2020 May 29;17(11) [FREE Full text] [doi: 10.3390/ijerph17113857] [Medline: 32485859]
21. Leung GM, Lam T, Ho L, Ho S, Chan BHY, Wong IOL, et al. The impact of community psychological responses on outbreak control for severe acute respiratory syndrome in Hong Kong. J Epidemiol Community Health 2003 Nov;57(11):857-863 [FREE Full text] [doi: 10.1136/jech.57.11.857] [Medline: 14600110]
22. El Ansari W, Labeeb S, Moseley L, Kotb S, El-Houfy A. Physical and psychological well-being of university students: survey of eleven faculties in Egypt. Int J Prev Med 2013 Mar;4(3):293-310 [FREE Full text] [Medline: 23626886]
23. Fan J, Liu X, Shao G, Qi J, Li Y, Pan W, et al. The epidemiology of reverse transmission of COVID-19 in Gansu Province, China. Travel Med Infect Dis 2020;37:101741 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101741] [Medline: 32407893]
24. McAlonan GM, Lee AM, Cheung V, Cheung C, Tsang KW, Sham PC, et al. Immediate and sustained psychological impact of an emerging infectious disease outbreak on health care workers. Can J Psychiatry 2007 Apr 23;52(4):241-247. [doi: 10.1177/070674370705200406] [Medline: 17500305]
25. Tam CWC, Pang EPF, Lam LCW, Chiu HFK. Severe acute respiratory syndrome (SARS) in Hong Kong in 2003: stress and psychological impact among frontline healthcare workers. Psychol Med 2004 Oct;34(7):1197-1204. [doi: 10.1017/s0033291704002247] [Medline: 15697046]
26. Grace SL, Hershenfield K, Robertson E, Stewart DE. The occupational and psychosocial impact of SARS on academic physicians in three affected hospitals. Psychosomatics 2005 Sep;46(5):385-391 [FREE Full text] [doi: 10.1176/appi.psy.46.5.385] [Medline: 16145182]
27. Zung WW. A rating instrument for anxiety disorders. Psychosomatics 1971 Nov;12(6):371-379. [doi: 10.1016/S0033-3182(71)71479-0]
28. Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965 Jan 01;12(1):63-70. [doi: 10.1001/archpsyc.1965.01720310065008] [Medline: 14221692]
29. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov 01;46(11):1006-1011. [doi: 10.1001/archpsyc.1989.01810110048007] [Medline: 2684084]
30. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
31. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5) [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
32. Li Z, Ge J, Yang M, Feng J, Qiao M, Jiang R, et al. Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. Brain Behav Immun 2020 Aug;88:916-919 [FREE Full text] [doi: 10.1016/j.bbi.2020.03.007] [Medline: 32169498]
33. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the United States: interview survey study. J Med Internet Res 2020 Sep 03;22(9):e21279 [FREE Full text] [doi: 10.2196/21279] [Medline: 32805704]
34. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020 Sep 17;22(9):e22817 [FREE Full text] [doi: 10.2196/22817] [Medline: 32897868]
35. Mira JJ, Cobos A, Martínez García O, Bueno Domínguez MJ, Astier-Peña MP, Pérez Pérez P, SARS-CoV-2 Second Victims Working Group. An acute stress scale for health care professionals caring for patients with COVID-19: validation study. JMIR Form Res 2021 Mar 09;5(3):e27107 [FREE Full text] [doi: 10.2196/27107] [Medline: 33687343]
36. Tan BY, Chew NW, Lee GK, Jing M, Goh Y, Yeo LL, et al. Psychological impact of the COVID-19 pandemic on health care workers in Singapore. Ann Intern Med 2020 Aug 18;173(4):317-320. [doi: 10.7326/m20-1083]
37. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
38. Kleiman EM, Yeager AL, Grove JL, Kellerman JK, Kim JS. Real-time mental health impact of the COVID-19 pandemic on college students: ecological momentary assessment study. JMIR Ment Health 2020 Dec 15;7(12):e24815 [FREE Full text] [doi: 10.2196/24815] [Medline: 33207308]
39. Copeland WE, McGinnis E, Bai Y, Adams Z, Nardone H, Devadanam V, et al. Impact of COVID-19 pandemic on college student mental health and wellness. J Am Acad Child Adolesc Psychiatry 2021 Jan;60(1):134-141.e2 [FREE Full text] [doi: 10.1016/j.jaac.2020.08.466] [Medline: 33091568]

XSL•FO
RenderX

40. Bolatov AK, Seisembekov TZ, Askarova AZ, Baikanova RK, Smailova DS, Fabbro E. Online-learning due to COVID-19 improved mental health among medical students. Med Sci Educ 2020 Nov 18;31(1):1-10 [FREE Full text] [doi: 10.1007/s40670-020-01165-y] [Medline: 33230424]
41. Li Y, Zhao J, Ma Z, McReynolds LS, Lin D, Chen Z, et al. Mental health among college students during the COVID-19 pandemic in China: a 2-wave longitudinal survey. J Affect Disord 2021 Feb 15;281:597-604. [doi: 10.1016/j.jad.2020.11.109] [Medline: 33257043]
42. Reznik A, Gritsenko V, Konstantinov V, Yehudai M, Bender S, Shilina I, et al. First and second wave COVID-19 fear impact: Israeli and Russian social work student fear, mental health and substance use. Int J Ment Health Addict 2021 Feb 02:1-8 [FREE Full text] [doi: 10.1007/s11469-020-00481-z] [Medline: 33551690]
43. Sun J, Li Z, Buys N, Storch EA. Correlates of comorbid depression, anxiety and helplessness with obsessive-compulsive disorder in Chinese adolescents. J Affect Disord 2015 Mar 15;174:31-37. [doi: 10.1016/j.jad.2014.11.004] [Medline: 25479051]
44. Velten J, Bieda A, Scholten S, Wannemüller A, Margraf J. Lifestyle choices and mental health: a longitudinal survey with German and Chinese students. BMC Public Health 2018 May 16;18(1):632 [FREE Full text] [doi: 10.1186/s12889-018-5526-2] [Medline: 29769115]
45. da Cunha J, Maselli LMF, Stern ACB, Spada C, Bydlowski SP. Impact of antiretroviral therapy on lipid metabolism of human immunodeficiency virus-infected patients: old and new drugs. World J Virol 2015 May 12;4(2):56-77 [FREE Full text] [doi: 10.5501/wjv.v4.i2.56] [Medline: 25964872]

Abbreviations

AIC: Akaike information criterion
OCD: obsessive-compulsive disorder
OR: odds ratio
SARS: severe acute respiratory syndrome
SAS: Self-Rating Anxiety Scale
SDS: Self-Rating Depression Scale
YBOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by C Basch; submitted 01.04.21; peer-reviewed by Z Yongjian, R Ho; comments to author 31.05.21; revised version received 20.06.21; accepted 20.06.21; published 21.07.21.

Please cite as:
Zhang X, Shi X, Wang Y, Jing H, Zhai Q, Li K, Zhao D, Zhong S, Song Y, Zhang F, Bao Y
Risk Factors of Psychological Responses of Chinese University Students During the COVID-19 Outbreak: Cross-sectional Web-Based Survey Study
J Med Internet Res 2021;23(7):e29312
URL: https://www.jmir.org/2021/7/e29312
doi:10.2196/29312
PMID:34156961

©Xudong Zhang, Xin Shi, Yang Wang, Huiquan Jing, Qingqing Zhai, Kunhang Li, Dan Zhao, Shiyu Zhong, Yuequn Song, Feng Zhang, Yijun Bao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 21.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Age and Social Disparities in the Use of Telemedicine During the COVID-19 Pandemic in Japan: Cross-sectional Study

Atsushi Miyawaki[1], MD, PhD; Takahiro Tabuchi[2], MD, PhD; Michael K Ong[3,4,5], MD, PhD; Yusuke Tsugawa[3,4], MPH, MD, PhD

[1]Department of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Cancer Control Center, Osaka International Cancer Institute, Osaka, Japan

[3]Division of General Internal Medicine and Health Services Research, UCLA David Geffen School of Medicine, Los Angeles, CA, United States

[4]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, CA, United States

[5]VA Center for the Study of Healthcare Innovation, Implementation and Policy, Los Angeles, CA, United States

Corresponding Author:
Atsushi Miyawaki, MD, PhD
Department of Public Health
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 1130033
Japan
Phone: 81 3 5841 3494
Email: amiyawaki-tky@umin.ac.jp

Abstract

Background: The use of telemedicine outpatient visits has increased dramatically during the COVID-19 pandemic in many countries. Although disparities in access to telemedicine by age and socioeconomic status (SES) have been well-documented, evidence is limited as to how these disparities changed during the COVID-19 pandemic. Moreover, the equity of patient access to telemedicine has been scarcely reported in Japan, despite the huge potential for telemedicine expansion.

Objective: We aimed to investigate changes due to age and SES disparities in telemedicine use during the COVID-19 pandemic in Japan.

Methods: Using data from a large internet survey conducted between August 25 and September 30, 2020, in Japan, we examined the associations of participant age and SES (educational attainment, urbanicity of residence, and income level) with their telemedicine use in the following two time periods during the pandemic: April 2020 and August-September 2020.

Results: Of the 24,526 participants aged 18 to 79 years (50.8% [n=12,446] women), the proportion of individuals who reported using telemedicine increased from 2.0% (n=497) in April 2020 to 4.7% (n=1159) in August-September 2020. After adjusting for potential confounders, younger individuals were more likely to use telemedicine than older individuals in April 2020. Although this pattern persisted in August-September 2020, we also observed a substantial increase in telemedicine use among individuals aged 70 to 79 years (adjusted rates, 0.2% in April 2020 vs 3.8% in August-September 2020; $P<.001$ after multiple comparisons). We found disparities in telemedicine use by SES in August-September 2020 that did not exist in April 2020. In August-September 2020, individuals with a university degree were more likely to use telemedicine than those with a high school diploma or less (adjusted rates, 6.6% vs 3.5%; $P<.001$). Individuals living in urban areas exhibited higher rates of telemedicine use than those living in rural areas only in August-September 2020 (adjusted rates, 5.2% vs 3.8%; $P<.001$). Disparities in telemedicine use by income level were not observed in either time period.

Conclusions: In general, younger individuals increased their use of telemedicine compared to older individuals during the pandemic, although individuals in their 70s also increased their use of telemedicine. Disparities in telemedicine use by educational attainment and urbanicity of residence widened during the COVID-19 pandemic.

(J Med Internet Res 2021;23(7):e27982) doi:10.2196/27982

KEYWORDS

telemedicine; telehealth; disparity; access to care; COVID-19; Japan

XSL•FO
RenderX

Introduction

The COVID-19 pandemic has led to a global public health crisis. More than 81 million cases have been reported worldwide as of the end of December 2020, with 1.8 million deaths from COVID-19 infection [1]. This pandemic has disrupted routine health services across countries, with in-person outpatient visits drastically decreasing in the United States [2,3] and Japan [4]. Telemedicine visits are partially replacing this decline in in-person visits [3,5], and they allow for medical care at a safe social distance in many countries [6]. Changes in the regulations related to the use of telemedicine, including higher reimbursement rates and less restrictive security requirements (eg, allowing non-Health Insurance Portability and Accountability Act–compliant modalities such as Facetime and Skype), facilitated increases in telemedicine usage [6,7].

Telemedicine has the potential to increase access to care for historically underserved populations who experience distance and transportation barriers to the more traditional face-to-face method of providing medical care [8]. Evidence suggests that telemedicine is effective for several types of medical interventions, including smoking cessation [9], psychiatry care [10,11], and management of chronic diseases, such as diabetes, heart disease, and chronic obstructive pulmonary disease [12,13]. Telemedicine can improve health outcomes even among older adults who face additional hurdles when using telemedicine due to sensory barriers and multiple comorbidities [14]. However, there have been concerns that the diffusion of telemedicine technology may have exacerbated social inequality in access to new health care technologies [15]. Racial/ethnic minorities, rural residents, and those with lower educational attainment have relatively limited access to information and communications technologies (ICTs) [16,17], and thus, they are less likely to take advantage of telemedicine. Older adults are also less prepared to use telemedicine than younger adults because of disabilities in hearing and speaking or inexperience with technology [16,18].

Disparities in telemedicine access by age and socioeconomic status (SES) have been well-documented during the COVID-19 pandemic. A study conducted at four clinics in San Francisco reported that the proportion of primary care visits with minority patients aged ≥65 years decreased after telemedicine implementation [19]. Other studies conducted in the United States have found that telemedicine during COVID-19 was concentrated among younger individuals living in urban areas with higher SES [20-23]. Yet, these findings show disparities only at one point in the COVID-19 pandemic, and evidence is limited as to how disparities in telemedicine use by age and SES changed as the need for telemedicine persisted during the COVID-19 pandemic.

Moreover, there is limited research on the sociodemographic patterns in telemedicine use outside the United States. Japan has universal health coverage through the social insurance scheme and is one of the countries with the highest life expectancy in the world [24]. Japan has a lower number of telemedicine users compared to the United States, Canada, and European countries [25]. Recently, mainly due to the impact of the COVID-19 pandemic, policymakers in Japan have been promoting the use of telemedicine. For example, the Japanese government approved the insurance coverage of telemedicine visits in 2018, and further promoted the use of telemedicine during the COVID-19 pandemic. Although the overall use of telemedicine has increased during the COVID-19 pandemic, given that it requires patients to be familiar with health care ICTs, it is possible that it has increased disparity gaps in access to and use of telemedicine between certain subpopulations (eg, old vs young populations). However, the equity of patient access to telemedicine has not been well-characterized in Japan.

To bridge this knowledge gap, we used data from a large internet survey conducted in Japan to examine how disparities in telemedicine use by age and SES may have changed during two time periods following the declaration of the pandemic: April 2020 and August-September 2020. According to the trend in confirmed COVID-19 cases reported in Japan, April 2020 corresponded to the early stage of the first wave of the COVID-19 epidemic (under the state of emergency), and August-September 2020 was during the second wave [1]. The aim of this study was to elucidate how disparities in telemedicine use by age and SES have changed during the pandemic.

Methods

Study Design, Setting, and Data Sources

We analyzed data from the *Japan "COVID-19 and Society" Internet Survey (JACSIS)* study [26], a cross-sectional, web-based, self-reported questionnaire survey administered by a major nationwide internet research agency with 2.2 million qualified panelists [27-29]. Selected individuals aged 15 to 79 years (n=224,389) were included in the JACSIS study using stratified random sampling based on gender, age, and prefecture category to represent the distribution of the general population in Japan in 2019 [30,31]. Individuals who agreed to participate in the survey accessed the designated website and responded to questionnaires asking about a wide range of socioeconomic, lifestyle, and health measures in the context of the COVID-19 pandemic. Questionnaires were distributed from August 25, 2020, until September 30, 2020 (hereafter, "August-September 2020"), when the predetermined target number of participants for each gender, age, and prefecture category was met, with an overall response rate of 12.5% (28,000/224,389). For this study, we used a subsample of adults aged 18 to 79 years (n=27,641). We sequentially excluded 2477 individuals showing unnatural or inconsistent responses using the algorithm we developed and 638 individuals whose zip code information was missing. This study was approved by the University of Tokyo (number 2020337NI).

Exposure Variables

The primary exposure variable was participant age. The secondary exposure variables were SES measures of educational attainment, urbanicity of residence, and household income level. Age was categorized into the following six groups: 18-29 (reference), 30-39, 40-49, 50-59, 60-69, and 70-79 years. Educational attainment was categorized into the following three groups: those who had a university degree or higher (reference; corresponding to the *International Standard Classification of*

XSL•FO
RenderX

Education [ISCED] [32] level 6-8), a college degree (ISCED level 5; colleges included junior, community, and vocational colleges), and a high school diploma or lower (ISCED level 0-4). Urbanicity of residence was dichotomized (urban [reference] vs rural). An urban area was defined as a densely inhabited district, which was determined according to the information from the 2015 Census of Japan linked to the seven-digit zip codes reported by the participants [33]; rural area was defined as a nonurban area. Income level was based on self-reported household income and categorized into four groups using the tertiles of household equivalent income ("high" = more than 4.3 million JPY [reference] [110 JPY ≒ 1 USD], "medium" = 2.5 to 4.3 million JPY, or "low" = less than 2.5 million JPY) and an indicator for those who refused to respond to this question.

Outcome Variables

Our outcome of interest was whether the participant used telemedicine, defined as the examination, diagnosis, and treatment of patients by physicians via information and communication devices that contain real-time visual and auditory information (according to the Ministry of Health, Labour, and Welfare of Japan) [34]. Telemedicine use was measured at the following two different time points during the COVID-19 pandemic in 2020: (1) April and (2) August-September (at the time of the survey). We asked the participants the following question: "Have you ever used telemedicine as a patient?" Participants were to select from one of the following three options: (1) "I have used it before April 2020," (2) "I used it for the first time after April 2020," and (3) "Never." We regarded the participants who chose option 1 as using telemedicine as of April and those who chose option 1 or 2 as using telemedicine as of August-September. We defined the use of telemedicine as those individuals who ever used telemedicine, rather than using the frequency of telemedicine use as the outcome variable, because only survey participants with health issues that require medical attention actually use telemedicine (even those individuals with access to telemedicine would not use telemedicine if they had no health issue during the study period). Moreover, the frequency of telemedicine was affected by not only patients' access to telemedicine, but also several other factors including the severity of illness of patients and the physicians' practice patterns (some physicians may see patients more often than others). A similar approach has been used in prior studies [35,36].

Adjustment Variables

We adjusted for the participants' sociodemographic and health-related characteristics. The sociodemographic characteristics included employment status (employer, self-employed, employee, and unemployed), marital status (married, never married, widowed, and separated), and household size (number of household members: 1, 2, 3, 4, and 5+). Health-related characteristics included smoking status (never, ever, and current smokers), self-rated health (excellent/good vs moderate/bad/very bad), walking disability (whether the person is experiencing difficulties in walking), and dummy variables for each of nine comorbidities (overweight [BMI ≥25 kg/m^2] and self-reported presence of eight conditions including hypertension, diabetes, asthma, coronary disease, stroke, chronic obstructive pulmonary disease, cancer, and psychological disorder). BMI was calculated by dividing self-reported body weight (kg) by self-reported body height squared (m^2).

Statistical Analysis

First, we described the sociodemographic and health-related characteristics of the participants. To account for the possibility that those who participated and responded to the internet-based survey might differ from the general population, we applied an inverse probability weighting (IPW) approach throughout the analyses [37]. The details of the calculation for IPW are described in Multimedia Appendix 1.

Second, we examined the association between age and rates of telemedicine use in April 2020 or August-September 2020. For each outcome, we constructed a weighted multivariable logistic regression model (IPW described above) that controlled for potential confounders (the other exposures [ie, educational attainment, urbanicity of residence, and income level] and sociodemographic/health-related characteristics). Standard errors were clustered at the prefecture level to account for the potential correlation of participants within the same prefecture. Japan consists of 47 prefectures, which are the country's first jurisdiction and administrative division levels. To calculate adjusted rates of telemedicine use, we employed marginal standardization (also known as predictive margins or margins of response). For each participant, we calculated predicted probabilities of telemedicine use with the exposure fixed at each category and then averaged over the distribution of the adjustment variables in our sample.

Third, to examine the adjusted change in telemedicine use from April 2020 to August-September 2020, we calculated the difference in the adjusted rates of telemedicine use between these two time points for each age group.

Finally, we tested if the adjusted changes in telemedicine use from April 2020 to August-September 2020 varied across age, by calculating the difference in differences. We also used a similar approach to examine the adjusted rates of telemedicine use and their changes by each SES aspect (educational attainment, urbanicity of residence, and income level).

We used the Benjamini-Hochberg method to account for the multiple comparisons across 15 (6+3+2+4) exposure categories (11 [5+2+1+3] categories for the difference in the adjusted differences), and reported both unadjusted and adjusted *P* values (adjusted *P* values <.05 were considered as statistically significant) [38,39]. All analyses were conducted using Stata version 15 (StataCorp LLC).

Secondary Analyses

We conducted sensitivity analyses. First, we additionally adjusted for indicator variables for each prefecture (prefecture fixed effects) to effectively compare individuals living in the same prefecture. Second, we additionally adjusted for four categorical variables representing availability of ICTs, including internet access at home and ownership of personal computers, smartphones, and tablet computers, to test whether online access

XSL•FO
RenderX

could explain the observed disparities in the rate of telemedicine use by age and SES.

Results

Characteristics of the Participants

A total of 24,526 participants (88.7% of the total adult subsample) were included in our analyses. There were 12,446 women (50.8%), and the mean age at the time of the survey was 50.1 (SD 16.6) years. Those who had a high school diploma or less were the most numerous, accounting for half of the participants (12,030/24,526, 49.1%), while one-third (7915/24,526, 32.3%) of the participants had a university degree or higher, and less than 20% (4581/24,526, 18.7%) had a college degree. Approximately 60% (14,666/24,526, 59.8%) of the participants lived in urban areas (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the participants.

Characteristics	Value (N=24,526)
Female, n (%)	12,446 (50.8%)
Age (years), mean (SD)	50.1 (16.6)
Educational attainment, n (%)	
University degree or higher	7915 (32.3%)
College degree[a]	4581 (18.7%)
High school diploma or lower	12,030 (49.1%)
Urbanicity of residence, n (%)	
Urban	14,666 (59.8%)
Rural	9560 (40.2%)
Income level, n (%)	
High	5458 (22.3%)
Medium	6814 (27.8%)
Low	7151 (29.2%)
Not answered	5103 (20.8%)
Employment status, n (%)	
Employer	1011 (4.1%)
Self-employed	1892 (7.7%)
Employee	12,623 (51.5%)
Unemployed	9000 (36.7%)
Marital status, n (%)	
Married	16,102 (65.7%)
Never married	5413 (22.1%)
Widowed	1606 (6.6%)
Separated	1403 (5.7%)
Household size, n (%)	
1	4152 (16.9%)
2	8386 (34.2%)
3	5547 (22.6%)
4	4124 (16.8%)
≥5	2317 (9.5%)
Smoking status, n (%)	
Never	12,186 (49.7%)
Ever	7423 (30.3%)
Current	4917 (20.0%)
Self-rated good health[b], n (%)	9839 (40.1%)
Walking disability, n (%)	3142 (12.8%)
Comorbidities, n (%)	
Overweight	5202 (21.2%)
Hypertension	5086 (20.7%)
Diabetes	1911 (7.8%)
Asthma	1450 (5.9%)

XSL•FO
RenderX

Characteristics	Value (N=24,526)
Coronary disease	923 (3.8%)
Stroke	459 (1.9%)
Chronic obstructive pulmonary disease	332 (1.4%)
Cancer	774 (3.2%)
Psychological disorder	1848 (7.5%)

[a]College includes junior, community, and vocational college.

[b]Self-rated good health was defined as very good or good using the 5-point Likert scale question, which asked if self-rated health status was very good, good, moderate, bad, or very bad.

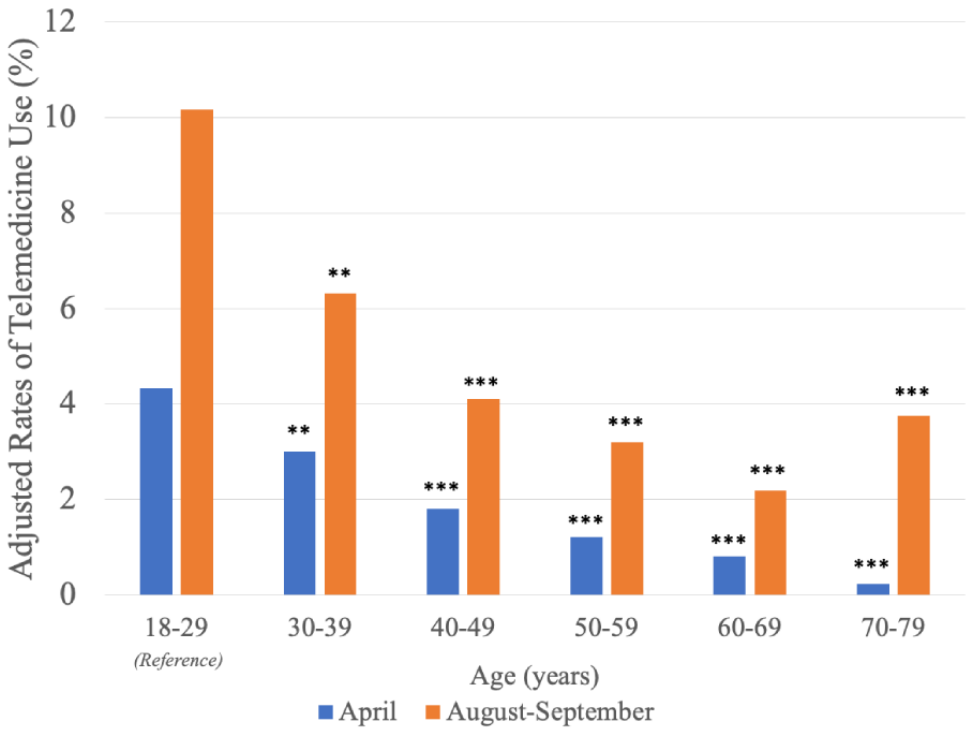
Overall Trend in the Rates of Telemedicine Use

Of the 24,526 participants aged 18 to 79 years (50.8% [n=12,446] women), the proportion of individuals who reported using telemedicine increased from 2.0% (n=497) in April 2020 to 4.7% (n=1159) in August-September 2020.

Adjusted Rates of Telemedicine Use by Age

After adjusting for potential confounders, younger individuals were more likely to use telemedicine than older individuals in April 2020 (Figure 1 and Multimedia Appendix 2). Although this pattern remained largely unchanged in August-September 2020, with participants aged 18 to 29 years exhibiting the largest increase in telemedicine use (adjusted rates, 4.3% in April vs 10.2% in August-September; adjusted difference, +5.8 percentage points; adjusted P<.001), we also observed a substantial increase in telemedicine use among participants aged 70 to 79 years (from 0.2% to 3.8%; +3.5 percentage points; P<.001) (Table 2). The increase in the adjusted rates of telemedicine use among participants aged 18 to 29 years was 3.5 percentage points larger compared to that among those aged 40 to 49 years (P=.04), 3.9 percentage points larger compared to that among those aged 50 to 59 years (P=.01), and 4.5 percentage points larger compared to that among those aged 60 to 69 years (P=.003). However, we found no evidence that the increase in the rates of telemedicine use differed between those aged 18 to 29 years and 70 to 79 years (P=.19).

Figure 1. Adjusted rates of telemedicine use in April 2020 and August-September 2020 by age. Telemedicine use was defined as "ever use" of telemedicine at a given time point (April or August-September 2020). Rates of telemedicine use were adjusted for other exposures and sociodemographic/health-related characteristics. We statistically compared the adjusted rates of telemedicine by age groups (reference: age 18-29 years) in April 2020 and in August-September 2020. **Adjusted P<.01, ***Adjusted P<.001.



XSL•FO
RenderX

Table 2. Difference in the adjusted rates of telemedicine use between April 2020 and August-September 2020 by age.

Age group (years)	Number of participants	Adjusted rate, %		Difference[a], % (95% CI)	*P* value[b]	Difference in differences[a] (95% CI)	*P* value[b]
		April 2020	August-September 2020				
18-29	3388	4.3	10.2	5.8 (3.0 to 8.7)	<.001	Reference	N/A[c]
30-39	3784	3.0	6.3	3.3 (1.7 to 4.9)	<.001	−2.5 (−5.8 to 0.7)	.19
40-49	4883	1.8	4.1	2.3 (1.2 to 3.3)	<.001	−3.5 (−6.6 to −0.5)	.04
50-59	4278	1.2	3.2	2.0 (0.9 to 3.1)	<.001	−3.9 (−6.9 to −0.8)	.01
60-69	4286	0.8	2.2	1.4 (0.4 to 2.4)	.004	−4.5 (−7.5 to −1.4)	.003
70-79	3907	0.2	3.8	3.5 (2.2 to 4.8)	<.001	−2.3 (−5.5 to 0.8)	.19

[a]We calculated the differences in the adjusted rates of telemedicine use between April 2020 and August-September 2020 for each age group. Then, we examined how the difference in the rates of telemedicine use between the two time points varied by age (difference in differences). The analyses were weighted to account for selection in an internet survey. For each analysis, standard errors were clustered at the prefecture level.

[b]The *P* values were adjusted post hoc to account for multiple comparisons with the use of the Benjamini-Hochberg method.

[c]N/A: not applicable.

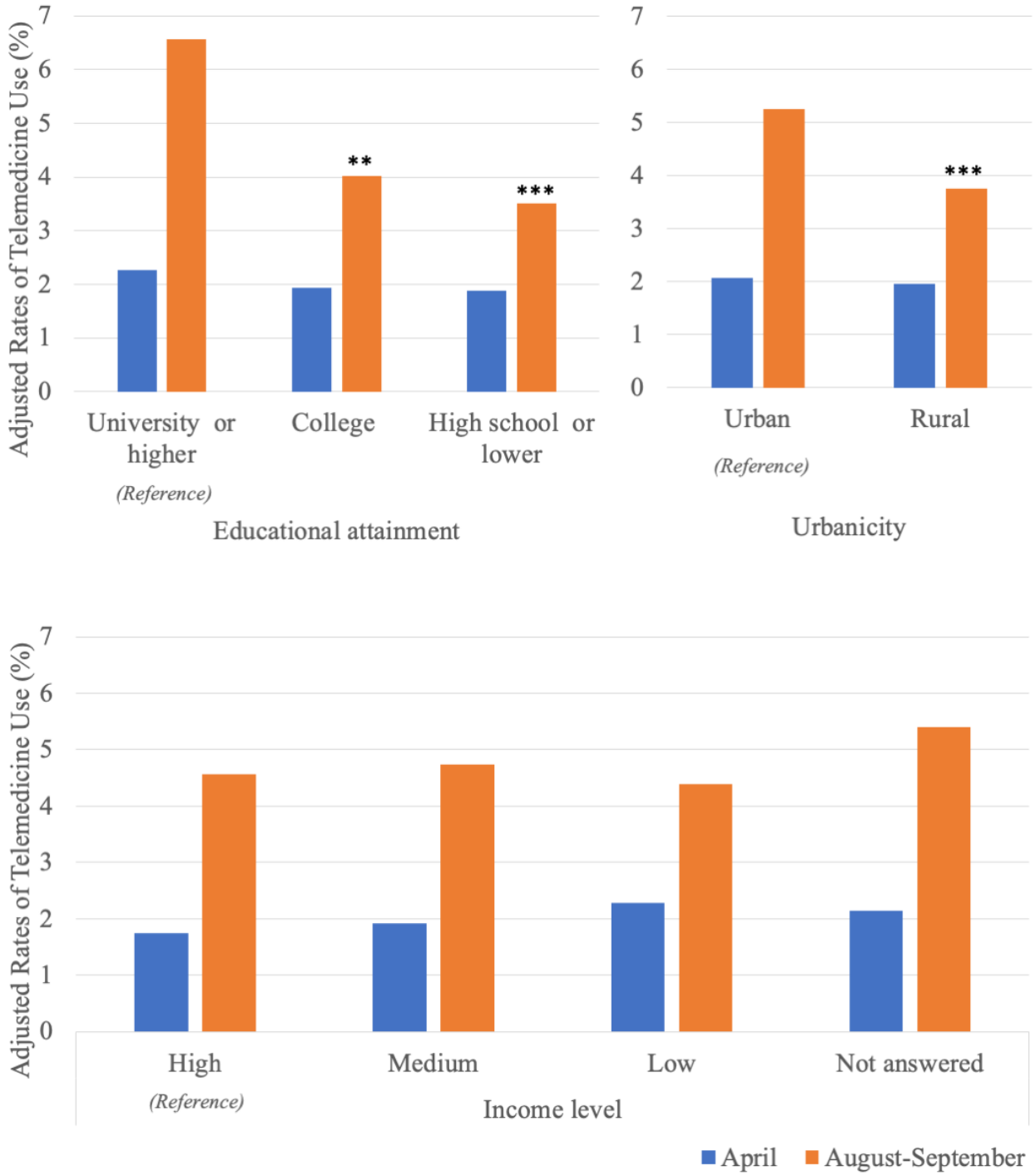
Adjusted Rates of Telemedicine Use by SES Measures

Educational Attainment

We found no evidence that the adjusted rates of telemedicine use differed by educational attainment in April 2020 (Figure 2). In contrast, we found disparities in telemedicine use by SES in August-September 2020. Participants with a university degree or higher were more likely to use telemedicine than those with a college degree (adjusted rates, 6.6% vs 4.0%; *P*=.006) or high school diploma or less (adjusted rates, 6.6% vs 3.5%; *P*<.001) (Multimedia Appendix 2). The increase in the adjusted rates of telemedicine use among participants with a university degree or higher was 2.2 percentage points larger compared to that among those with a college degree (*P*=.01) and 2.7 percentage points larger compared to that among those with a high school diploma or less (*P*=.003) (Table 3).

XSL•FO
RenderX

Figure 2. Adjusted rates of telemedicine use in April 2020 and August-September 2020 by socioeconomic status measures. Telemedicine use was defined as "ever use" of telemedicine at a given time point (April or August-September 2020). Rates of telemedicine use were adjusted for other exposures and sociodemographic/health-related characteristics. We statistically compared the adjusted rates of telemedicine by educational attainment (reference: having a university degree or more), urbanicity of residence (reference: living in urban areas), or income level (reference: high income) in April 2020 and in August-September 2020. **Adjusted *P*<.01, ***Adjusted *P*<.001.



XSL•FO
RenderX

Table 3. Difference in the adjusted rates of telemedicine use between April 2020 and August-September 2020 by socioeconomic status measures.

Measure	Number of participants	Adjusted rate, %		Difference[a], % (95% CI)	*P* value[b]	Difference in differences[a] (95% CI)	*P* value[b]
		April 2020	August-September 2020				
Educational attainment							
University or higher	7915	2.3	6.6	4.3 (2.7 to 5.9)	<.001	Reference	N/A[c]
College	4581	1.9	4.0	2.1 (1.3 to 2.9)	<.001	−2.2 (−4.0 to −0.4)	.01
High school or lower	12,030	1.9	3.5	1.6 (0.9 to 2.4)	<.001	−2.7 (−4.4 to −0.9)	.003
Urbanicity of residence							
Urban	14,666	2.1	5.2	3.2 (2.4 to 3.9)	<.001	Reference	N/A
Rural	9860	2.0	3.8	1.8 (1.2 to 2.4)	<.001	−1.4 (−2.3 to −0.4)	.004
Income level							
High	5458	1.7	4.6	2.8 (1.8 to 3.8)	<.001	Reference	N/A
Medium	6814	1.9	4.7	2.8 (1.5 to 4.2)	<.001	0 (−1.6 to 1.7)	.99
Low	7151	2.3	4.4	2.1 (1.3 to 2.9)	<.001	−0.7 (−2.0 to 0.6)	.34
Not answered	5103	2.1	5.4	3.2 (1.1 to 5.4)	.003	0.4 (−2.0 to 2.8)	.79

[a]We calculated the differences in the adjusted rates of telemedicine use between April 2020 and August-September 2020 for each socioeconomic status measure (educational attainment, urbanicity of residence, or income level). Then, we examined how the difference in the rates of telemedicine use between the two time points varied by educational attainment, urbanicity of residence, or income level (difference in differences). The analyses were weighted to account for selection in an internet survey. For each analysis, standard errors were clustered at the prefecture level.

[b]The *P* values were adjusted post hoc to account for multiple comparisons with the use of the Benjamini-Hochberg method.

[c]N/A: not applicable.

Urbanicity of Residence

We found no evidence that the adjusted rates of telemedicine use differed by urbanicity of residence in April 2020 (Figure 2). Participants living in urban areas exhibited higher rates of telemedicine use than those living in rural areas only in August-September 2020 (adjusted rates, 5.2% vs 3.8%; *P*<.001) (Multimedia Appendix 2). The increase in the adjusted rates of telemedicine use among participants living in urban areas was 1.4 percentage points higher compared to that among those living in rural areas (*P*=.004) (Table 3).

Income Level

We found no evidence that the adjusted rates of telemedicine use differed by income level in April or August-September 2020 (Figure 2). We also found no evidence that the increase in the rates of telemedicine use varied by income level (Table 3).

Secondary Analyses

Our findings were qualitatively unaffected by additional adjustment for prefecture fixed effects (Multimedia Appendix 3). When we additionally adjusted for the variables representing the availability of ICTs, our findings of increased disparities by SES (ie, university or higher vs high school or lower and urban vs rural) were largely unaffected (Multimedia Appendix 4). However, we no longer found any evidence that the increase in the rates of telemedicine use varied between young individuals (aged 18-29 years) and middle-aged individuals (aged 40-49, 50-59, or 60-69 years), indicating that the trends of the disparities in telemedicine use by age were partly driven by differences in online access.

Discussion

Principal Findings

Using data from a large nationwide internet survey conducted in Japan, we found that younger individuals increased the use of telemedicine more than older individuals (leading to wider disparities by age) during the COVID-19 pandemic, although individuals in their 70s also increased the use of telemedicine. Disparities in telemedicine use by educational attainment and urbanicity of residence also increased during the pandemic, whereas we found no evidence that disparities by income level changed.

Younger people were already more likely to use telemedicine as of April 2020; this disparity by age further increased as of August-September 2020 among adults aged 69 years or younger. However, telemedicine use also increased among older adults aged 70 to 79 years, narrowing the gap in telemedicine use among this population. In April 2020, there was no evidence that telemedicine use varied across educational attainment and urbanicity of residence, but by August-September 2020, better educated individuals living in urban areas were more likely to use telemedicine. In contrast, we did not find significant disparities by income level in either April or August-September 2020. Taken together, these findings suggest that improved access to telemedicine during the pandemic may have penetrated unevenly in the population, leaving behind some

XSL•FO
RenderX

socioeconomically disadvantaged populations. In contrast, the growing use of telemedicine among older adults was reassuring, indicating that this population of people, who have high health care needs but are often unfamiliar with ICTs [16], benefitted from improved access to telemedicine.

There may be several mechanisms through which the disparities in telemedicine use due to age have widened, especially among individuals aged 69 years or younger. First, younger people may have been more familiar with ICTs, and therefore, faced lower psychological or technological hurdles to initiating telemedicine usage when access to it improved in response to the COVID-19 pandemic in Japan [16]. This explanation is supported by our findings showing no evidence of disparities in telemedicine use due to age once we adjusted for the indicators of ICT availability. Second, the health care needs of young patients may be milder than those of older patients, making young patients more suited to the use of telemedicine. For example, young patients may have fewer comorbidities and seek care for milder conditions compared to middle-aged or older adults. Young patients also may be less likely to need blood tests or diagnostic imaging, which require patients to physically visit a health care facility [40,41]. Finally, younger individuals may be less likely to have disabilities and sensory barriers that could be hurdles to telemedicine use [14].

We also found that individuals aged 70 to 79 years experienced a large increase in telemedicine use. Given that these older adults are in the highest-risk age group for COVID-19 infection (most likely to experience severe life-threatening conditions when infected) [42], they might be incentivized to use telemedicine to avoid visiting health care facilities, which might lead to COVID-19 infection.

The limited increase in telemedicine use for individuals with lower academic attainment may be due to their lower digital literacy, limited access to ICTs, and less flexible work schedules [43], all of which could be barriers to using telemedicine [16]. The limited penetration of telemedicine use in rural areas may be explained by inadequate internet access in rural areas [16] and the fact that there were fewer medical institutions offering telemedicine in rural areas [44]. Given that the number of COVID-19 infections was generally higher in urban areas compared with rural areas, another potential explanation for the relative expansion in telemedicine use in urban areas may be that patients living in urban areas were more incentivized to avoid face-to-face encounters. However, we found that the inclusion of prefecture fixed effects (effectively comparing urban vs rural areas within the same prefecture) did not qualitatively change our results, indicating that this does not fully explain the disparities in telemedicine use between residents of urban and rural areas. It was reassuring that we found no substantial disparities in access to telemedicine by income level throughout the study period. Our findings may indicate that, at least in Japan, financial barriers have minimally influenced the disparities in telemedicine use.

Comparison With Prior Work

Our findings add to the body of work investigating the impact of age and SES on telemedicine use during the COVID-19 pandemic. Studies that were conducted prior to the expansion of telemedicine use in response to the COVID-19 pandemic have reported that younger patients are more likely to use telemedicine, whereas these studies showed mixed results regarding the disparities by area income level and urbanicity of residence. Studies that were conducted after the COVID-19 pandemic found that telemedicine use overall [41,45,46] and for geriatric care [21], primary care [19,22,47], or otolaryngological care [23] was concentrated among younger individuals living in urban high-income areas. Yet, the timeframe used in these studies represented only one early point in the COVID-19 pandemic, and the studies did not focus on the change over time of age and social disparities in telemedicine use. These previous studies used area income level as an indicator of income level, as opposed to individual income level used in our research. The inconsistent results for income level may partly be attributable to this difference. More importantly, to our knowledge, our study is the first to show that telemedicine use increased substantially among individuals aged 70 years or older, reducing telemedicine disparities among the population at the highest risk for COVID-19 infection.

Limitations

Our study has some limitations. First, as with any observational study, we could not fully account for unmeasured confounders. Our study also was unable to identify the exact mechanisms of the association between age or SES and the increase in the rates of telemedicine use. Second, due to a limitation in our data, we were unable to identify whether telemedicine use represented telephone visits or virtual visits. Third, there is potential for recall bias; younger individuals with higher SES might be more likely to recall and report telemedicine use. Fourth, we could not identify the clinical conditions for which patients received care via telemedicine. The reports submitted to the government by medical institutions offering telemedicine indicated that the most common conditions for which patients used telemedicine between July and September 2020 were acute relatively mild conditions, such as upper respiratory tract infection, fever, bronchitis, and rhinitis [44]. However, evidence suggests that telemedicine is suitable for psychiatric care [10,11] and management of chronic conditions [12,13]. This gap may be due to patients avoiding clinic visits with signs of infection and shifting to telemedicine, or the Japanese authorities' restriction on the prescription of psychiatric drugs in telemedicine [44]. The patterns of telemedicine use by age and SES may change as the profiles of conditions treated via telemedicine increase in the future. Finally, because our study sample was collected through an internet-based survey, our findings may not be generalizable to a population with limited access to and/or literacy of the internet. However, we used weighted analyses to minimize the difference in demographics, SES, and health-related characteristics between respondents of the current internet survey and the nationally representative survey, thus approximating our estimates to national estimates.

Conclusions

Using a large-scale nationwide internet survey in Japan, we found that younger individuals were generally more likely to increase telemedicine use than older individuals during the pandemic, although individuals in their 70s exceptionally gained

XSL•FO
RenderX

access to telemedicine. Disparities in telemedicine use by educational attainment and urbanicity of residence increased during the COVID-19 pandemic. These findings indicate that the current telemedicine expansion may be leaving a portion of the socioeconomically disadvantaged population behind, and suggest further need for policy efforts to achieve equal access to health care.

Acknowledgments

Data collection and statistical analyses in this work were funded by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grants (18H03062 to TT). AM was supported by JSPS KAKENHI Grants (20K18956) and the Social Science Research Council (grant number is not applicable) for other work not related to this study. YT was supported by the National Institutes of Health (NIH)/National Institute on Minority Health and Health Disparities Grant R01MD013913 and NIH/National Institute of Aging Grant R01AG068633 for other work not related to this study. The data from the JACSIS study were collected with the support of JSPS KAKENHI Grants (grant numbers 18H03062, 17H03589, 19K10671, 19K19439, 19K10446, and 18H03107), the University of Tsukuba Research Support Program (grant number is not applicable), and Health Labour Sciences Research Grants (19FA1005). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funders. The funders were not involved with the collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. All authors had full access to all the data in the study and had final responsibility for the decision to submit the paper.

Authors' Contributions

AM had full access to the data in the study and takes responsibility for the accuracy and integrity of the data and its analyses. Study concept and design: All authors. Acquisition, analysis, or interpretation of data: All authors. Drafting of the manuscript: AM and YT. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: AM, TT, and YT. Administrative, technical, or material support: YT. Study supervision: YT.

Conflicts of Interest

MKO receives royalties from UpToDate for its telemedicine section. The other authors have no conflicts to declare.

Multimedia Appendix 1
Inverse probability weighting.
[DOCX File , 22 KB - jmir_v23i7e27982_app1.docx]

Multimedia Appendix 2
Adjusted odds ratios of telemedicine for age and socioeconomic status measures.
[DOCX File , 21 KB - jmir_v23i7e27982_app2.docx]

Multimedia Appendix 3
Difference in adjusted rates of telemedicine use between April 2020 and August-September 2020 by age and socioeconomic status measures, additionally adjusting for prefecture fixed effects.
[DOCX File , 21 KB - jmir_v23i7e27982_app3.docx]

Multimedia Appendix 4
Difference in adjusted rates of telemedicine use between April 2020 and August-September 2020 by age and socioeconomic status measures, additionally adjusting for availability of information and communication technologies.
[DOCX File , 22 KB - jmir_v23i7e27982_app4.docx]

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2021-06-01]
2. Mehrotra A, Chernew M, Linetsky D, Hatch H, Cutler D, Schneider E. The Impact of the COVID-19 Pandemic on Outpatient Care: Visits Return to Prepandemic Levels, but Not for All Providers and Patients. The Commonwealth Fund. 2020. URL: https://www.commonwealthfund.org/publications/2020/oct/impact-covid-19-pandemic-outpatient-care-visits-return-prepandemic-levels [accessed 2021-06-01]
3. Baum A, Kaboli PJ, Schwartz MD. Reduced In-Person and Increased Telehealth Outpatient Visits During the COVID-19 Pandemic. Ann Intern Med 2021 Jan;174(1):129-131 [FREE Full text] [doi: 10.7326/M20-3026] [Medline: 32776780]
4. Hospital Report (August 2020). Ministry of Health Labour and Welfare. 2020. URL: https://www.mhlw.go.jp/toukei/list/80-1.html [accessed 2020-12-31]

XSL•FO
RenderX

5. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in Outpatient Care Delivery and Telemedicine During the COVID-19 Pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
6. Hassan A, Mari Z, Gatto EM, Cardozo A, Youn J, Okubadejo N, International Telemedicine Study Group. Global Survey on Telemedicine Utilization for Movement Disorders During the COVID-19 Pandemic. Mov Disord 2020 Oct 24;35(10):1701-1711 [FREE Full text] [doi: 10.1002/mds.28284] [Medline: 32833273]
7. Verma S. Early Impact Of CMS Expansion Of Medicare Telehealth During COVID-19. Health Affairs Blog. 2020 Jul 15. URL: https://www.healthaffairs.org/do/10.1377/hblog20200715.454789/full [accessed 2021-06-01]
8. WHO Global Observatory for eHealth. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. Geneva, Switzerland: World Health Organization; 2010. URL: https://apps.who.int/iris/handle/10665/44497
9. Myung S, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009 May 25;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]
10. Barak A, Hen L, Boniel-Nissim M, Shapira N. A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. Journal of Technology in Human Services 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
11. Hyler SE, Gangure DP, Batchelder ST. Can telepsychiatry replace in-person psychiatric assessments? A review and meta-analysis of comparison studies. CNS Spectr 2005 May 07;10(5):403-413. [doi: 10.1017/s109285290002277x] [Medline: 15858458]
12. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: A systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
13. Polisena J, Coyle D, Coyle K, McGill S. Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations. Int J Technol Assess Health Care 2009 Jul;25(3):339-349. [doi: 10.1017/S0266462309990201] [Medline: 19619353]
14. Batsis JA, DiMilia PR, Seo LM, Fortuna KL, Kennedy MA, Blunt HB, et al. Effectiveness of Ambulatory Telemedicine Care in Older Adults: A Systematic Review. J Am Geriatr Soc 2019 Aug;67(8):1737-1749 [FREE Full text] [doi: 10.1111/jgs.15959] [Medline: 31066916]
15. Timmermans S, Kaufman R. Technologies and Health Inequities. Annu. Rev. Sociol 2020 Jul 30;46(1):583-602. [doi: 10.1146/annurev-soc-121919-054802]
16. Internet/Broadband Fact Sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-06-01]
17. Nishida T, Pick JB, Sarkar A. Japan s prefectural digital divide: A multivariate and spatial analysis. Telecommunications Policy 2014 Dec;38(11):992-1010. [doi: 10.1016/j.telpol.2014.05.004]
18. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing Telemedicine Unreadiness Among Older Adults in the United States During the COVID-19 Pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
19. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst: Innovations in Care Delivery. 2020 May 04. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123 [accessed 2021-06-01]
20. Jaffe DH, Lee L, Huynh S, Haskell TP. Health Inequalities in the Use of Telehealth in the United States in the Lens of COVID-19. Popul Health Manag 2020 Oct;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
21. Rivera V, Aldridge MD, Ornstein K, Moody KA, Chun A. RESEARCHRacial and Socioeconomic Disparities in Access to Telehealth. J Am Geriatr Soc 2021 Jan;69(1):44-45. [doi: 10.1111/jgs.16904] [Medline: 33075143]
22. Pierce RP, Stevermer JJ. Disparities in use of telehealth at the onset of the COVID-19 public health emergency. J Telemed Telecare 2020 Oct 21:1357633X2096389. [doi: 10.1177/1357633x20963893]
23. Darrat I, Tam S, Boulis M, Williams AM. Socioeconomic Disparities in Patient Use of Telehealth During the Coronavirus Disease 2019 Surge. JAMA Otolaryngol Head Neck Surg 2021 Mar 01;147(3):287-295. [doi: 10.1001/jamaoto.2020.5161] [Medline: 33443539]
24. Ikeda N, Saito E, Kondo N, Inoue M, Ikeda S, Satoh T, et al. What has made the population of Japan healthy? The Lancet 2011 Sep;378(9796):1094-1105. [doi: 10.1016/s0140-6736(11)61055-6]
25. Market study on telemedicine. European Commission. 2018. URL: https://ec.europa.eu/health/sites/default/files/ehealth/docs/2018_provision_marketstudy_telemedicine_en.pdf [accessed 2021-06-05]
26. Ikeda T, Igarashi A, Odani S, Murakami M, Tabuchi T. Health-Related Quality of Life during COVID-19 Pandemic: Assessing Impacts of Job Loss and Financial Support Programs in Japan. Appl Res Qual Life 2021 Jan 30:1-17 [FREE Full text] [doi: 10.1007/s11482-021-09918-6] [Medline: 33552309]
27. About Us. Rakuten Insight. URL: https://insight.rakuten.co.jp/en/aboutus.html [accessed 2021-06-01]

XSL•FO
RenderX

28. Tabuchi T, Kiyohara K, Hoshino T, Bekki K, Inaba Y, Kunugita N. Awareness and use of electronic cigarettes and heat-not-burn tobacco products in Japan. Addiction 2016 Apr;111(4):706-713. [doi: 10.1111/add.13231] [Medline: 26566956]
29. Tabuchi T, Gallus S, Shinozaki T, Nakaya T, Kunugita N, Colwell B. Heat-not-burn tobacco product use in Japan: its prevalence, predictors and perceived symptoms from exposure to secondhand heat-not-burn tobacco aerosol. Tob Control 2018 Jul;27(e1):e25-e33 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053947] [Medline: 29248896]
30. Miyawaki A, Tabuchi T, Tomata Y, Tsugawa Y. Association between participation in the government subsidy programme for domestic travel and symptoms indicative of COVID-19 infection in Japan: cross-sectional study. BMJ Open 2021 Apr 13;11(4):e049069 [FREE Full text] [doi: 10.1136/bmjopen-2021-049069] [Medline: 33849861]
31. Gotanda H, Miyawaki A, Tabuchi T, Tsugawa Y. Association Between Trust in Government and Practice of Preventive Measures During the COVID-19 Pandemic in Japan. J Gen Intern Med 2021 Jun 22:1-7 [FREE Full text] [doi: 10.1007/s11606-021-06959-3] [Medline: 34159544]
32. International Standard Classification of Education: ISCED. Montreal, Quebec: UNESCO Institute for Statistics; 2012. URL: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf
33. Nakaya T, Honjo K, Hanibuchi T, Ikeda A, Iso H, Inoue M, Japan Public Health Center-based Prospective Study Group. Associations of all-cause mortality with census-based neighbourhood deprivation and population density in Japan: a multilevel survival analysis. PLoS One 2014;9(6):e97802 [FREE Full text] [doi: 10.1371/journal.pone.0097802] [Medline: 24905731]
34. Guidelines for the Appropriate Use of Telemedicine. Ministry of Health Labour and Welfare. 2019. URL: https://www.mhlw.go.jp/content/000534254.pdf [accessed 2021-06-08]
35. Belot M, Choi S, Jamison J, Papageorge N, Tripodi E, van den Broek-Altenburg E. Six-country survey on COVID-19. IZA – Institute of Labor Economics. 2020 May. URL: http://ftp.iza.org/dp13230.pdf [accessed 2021-06-01]
36. Mehta P, Braskett M, Tam JS, Espinoza J. Reimbursement patterns and user experiences in pediatric allergy home telehealth. Ann Allergy Asthma Immunol 2020 Dec;125(6):693-698.e1 [FREE Full text] [doi: 10.1016/j.anai.2020.06.003] [Medline: 32526378]
37. Schonlau M, van Soest A, Kapteyn A, Couper M. Selection Bias in Web Surveys and the Use of Propensity Scores. Sociological Methods & Research 2009 Feb 01;37(3):291-318. [doi: 10.1177/0049124108327128]
38. Sentilhes L, Winer N, Azria E, Sénat M, Le Ray C, Vardon D, et al. Tranexamic Acid for the Prevention of Blood Loss after Vaginal Delivery. N Engl J Med 2018 Aug 23;379(8):731-742. [doi: 10.1056/nejmoa1800942]
39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
40. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and Content of Primary Care Office-Based vs Telemedicine Care Visits During the COVID-19 Pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]
41. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation In Telemedicine Use And Outpatient Care During The COVID-19 Pandemic In The United States. Health Aff (Millwood) 2021 Feb 01;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
42. Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. Bull. World Health Organ 2020 Oct 14;99(1):19-33F. [doi: 10.2471/blt.20.265892]
43. Masso M. Determinants of employee work schedule and method control. Economic and Industrial Democracy 2012 Jul 09;34(3):451-469. [doi: 10.1177/0143831x12451348]
44. Summary of telemedicine use from July to September 2020. Ministry of Health Labour and Welfare. 2020. URL: https://www.mhlw.go.jp/content/10803000/000690548.pdf [accessed 2021-06-01]
45. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient Characteristics Associated With Telemedicine Access for Primary and Specialty Ambulatory Care During the COVID-19 Pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
46. Kakani P, Sorensen A, Quinton JK, Han M, Ong MK, Kamdar N, et al. Patient Characteristics Associated with Telemedicine Use at a Large Academic Health System Before and After COVID-19. J Gen Intern Med 2021 Apr;36(4):1166-1168 [FREE Full text] [doi: 10.1007/s11606-020-06544-0] [Medline: 33506387]
47. Schweiberger K, Hoberman A, Iagnemma J, Schoemer P, Squire J, Taormina J, et al. Practice-Level Variation in Telemedicine Use in a Pediatric Primary Care Network During the COVID-19 Pandemic: Retrospective Analysis and Survey Study. J Med Internet Res 2020 Dec 18;22(12):e24345 [FREE Full text] [doi: 10.2196/24345] [Medline: 33290244]

Abbreviations

ICT: information and communications technology
IPW: inverse probability weighting
ISCED: International Standard Classification of Education

XSL•FO
RenderX

JACSIS: Japan “COVID-19 and Society” Internet Survey
SES: socioeconomic status

Edited by C Basch; submitted 21.02.21; peer-reviewed by C Emezue, M Hamidzadeh; comments to author 31.05.21; revised version received 08.06.21; accepted 23.06.21; published 23.07.21.

Please cite as:
Miyawaki A, Tabuchi T, Ong MK, Tsugawa Y
Age and Social Disparities in the Use of Telemedicine During the COVID-19 Pandemic in Japan: Cross-sectional Study
J Med Internet Res 2021;23(7):e27982
URL: https://www.jmir.org/2021/7/e27982
doi:10.2196/27982
PMID:34259641

©Atsushi Miyawaki, Takahiro Tabuchi, Michael K Ong, Yusuke Tsugawa. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 23.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of a Parsimonious COVID-19 Outbreak Prediction Model: Heuristic Modeling Approach Using Publicly Available Data Sets

Agrayan K Gupta[1,2]; Shaun J Grannis[2,3], MD, MS; Suranga N Kasthurirathne[2,3], PhD

[1]Indiana University, Bloomington, IN, United States

[2]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

[3]School of Medicine, Indiana University, Indianapolis, IN, United States

Corresponding Author:
Agrayan K Gupta
Indiana University
107 S Indiana Ave
Bloomington, IN, 47405
United States
Phone: 1 812 855 4848
Email: aggupta@iu.edu

Abstract

Background: The COVID-19 pandemic has changed public health policies and human and community behaviors through lockdowns and mandates. Governments are rapidly evolving policies to increase hospital capacity and supply personal protective equipment and other equipment to mitigate disease spread in affected regions. Current models that predict COVID-19 case counts and spread are complex by nature and offer limited explainability and generalizability. This has highlighted the need for accurate and robust outbreak prediction models that balance model parsimony and performance.

Objective: We sought to leverage readily accessible data sets extracted from multiple states to train and evaluate a parsimonious predictive model capable of identifying county-level risk of COVID-19 outbreaks on a day-to-day basis.

Methods: Our modeling approach leveraged the following data inputs: COVID-19 case counts per county per day and county populations. We developed an outbreak gold standard across California, Indiana, and Iowa. The model utilized a per capita running 7-day sum of the case counts per county per day and the mean cumulative case count to develop baseline values. The model was trained with data recorded between March 1 and August 31, 2020, and tested on data recorded between September 1 and October 31, 2020.

Results: The model reported sensitivities of 81%, 92%, and 90% for California, Indiana, and Iowa, respectively. The precision in each state was above 85% while specificity and accuracy scores were generally >95%.

Conclusions: Our parsimonious model provides a generalizable and simple alternative approach to outbreak prediction. This methodology can be applied to diverse regions to help state officials and hospitals with resource allocation and to guide risk management, community education, and mitigation strategies.

(J Med Internet Res 2021;23(7):e28812) doi:10.2196/28812

KEYWORDS

coronavirus; COVID-19; emerging outbreak; modeling disease outbreak; precision public health; predictive modeling

Introduction

Background

The COVID-19 pandemic has impacted the health and well-being of individuals, communities, and economies worldwide at a hitherto unprecedented scale [1-3]. On March 11, 2020, the World Health Organization declared COVID-19 a pandemic with over 118,000 confirmed cases and 4291 deaths in over 114 countries [4]. To date, the pandemic has resulted in over 170 million confirmed cases, with over 3.5 million deaths globally [5]. In the United States, 33 million people have had COVID-19 and more than 600,000 lives have been lost [5].

At the height of the pandemic, waves of viral outbreaks placed health systems across the globe under extended strain, leading

XSL•FO
RenderX

to shortages in hospital beds, personal protective equipment, and health care personnel, which caused significant disruptions to health care delivery and loss of life [2,6]. Experts have estimated the cumulative financial costs of the COVID-19 pandemic related to lost output and health reduction at US $16 trillion, or approximately 90% of the annual gross domestic product of the United States [7].

In contrast with historical pandemics, the availability of public and population health information systems has enabled researchers to collaborate on many research activities in response to COVID-19 [8,9]. Since the onset of the pandemic, data scientists have collaborated with governmental organizations to create various public-facing COVID-19 dashboards that provide easy access to descriptive statistics and other metrics [10,11]. Information on COVID-19–related mortality, utilization of health care resources, and recovery has been crucial in increasing situational awareness to inform ongoing pandemic response efforts across communities [12,13].

Most recently, COVID-19 infection rates have started to decrease in response to increased vaccination and efforts in public education [14,15]. To date, 40% of the population of the United States is fully vaccinated [16]. These improvements have led to an interest in relaxing or revoking various restrictions enforced at the state and county levels. While important to the well-being of both communities and economies, such decisions may be dangerous if undertaken without adequate preplanning and awareness of potential risks. As such, effective identification of potential outbreaks of COVID-19 offers the ability to inform decision-makers across governmental and public health sectors on how to resume normal day-to-day activities in their communities and deploy limited human and treatment resources to where they are most needed [17].

Previous studies have demonstrated the potential to apply analytical models to identify potential outbreaks in response to other diseases [18]. However, these methods rely on large, complex data sets extracted from a specific health system or region [19-21]. Such data sets may be challenging and time-consuming to collect, leading to delays in generating timely predictions. Further, models trained using locale-specific data sets may not be generalizable across other locations [22], hindering the potential of reusing such models across other patient populations and regions. A variety of models are trained using complex algorithmic approaches such as neural networks and deep learning models. Such machine learning approaches may yield superior results but fail to achieve widespread acceptance [23] owing to challenges in explainability and interpretation [24].

In contrast, a less complex modeling approach that uses a subset of easily obtainable key elements widely captured across broad geographic regions may be less challenging to develop. Further, such models may also deliver adequate predictive performance without sacrificing explainability and interpretability. Such parsimonious models may also present less risk of overfitting on training data sets, thus allowing for greater generalizability [25].

Objective

We seek to leverage various readily accessible data sets extracted from multiple states to train and evaluate a parsimonious predictive model capable of identifying the county-level risk of COVID-19 outbreaks on a day-to-day basis.

Methods

Methods Overview

We selected 3 states for our efforts in COVID-19 outbreak prediction modeling: California, Indiana, and Iowa. These states were selected on the basis of geographical factors, governmental regulations, and availability of data sets for public use. For example, Indiana and Iowa are similar in the number of counties and total populations [26]. In contrast, California represented a more populous, urbanized state [26]. We also considered the general completeness of reporting, the quality of basic COVID-19 data sources, and the accuracy of state tracking systems [27].

Data Extraction and Cleaning

For each state, we extracted a variety of county-level data elements captured daily between March 1 and October 31, 2020. Data for Indiana were obtained from the Indiana State Department of Health, while data for Iowa and California were obtained from the New York Times web-based repository [5,28]. We selected March 1, 2020, as a start date as most states began collecting COVID-19 data at this time. October 31, 2020, marked the end of our analysis time period. Each data set was organized by county, state, and date reported using R [29]. Several errors or omissions in the data sets were addressed as follows: days with negative case counts were changed to 0 and a county labeled as “unknown” reported by Iowa and California were removed from further evaluation.

Preparation of a Gold Standard

We created a gold standard indicating if each county under study was in an outbreak on any particular day. A human expert reviewer created the gold standard by assigning an outbreak label (with responses of “yes” or “no”) to each county or date combination, considering the following criteria:

1. How do case counts trend in each county? Is there a general baseline of cases over time?
2. How large is the county’s population size (counties with more people report more cases)?
3. Duration of the outbreak to assign a binary indicator of “outbreak detected” or “outbreak not detected” to each day and county.

Based on our approach, a county could have multiple outbreaks over time. Outbreaks lasted a minimum of 3 days to account for testing lags as data were not always reported on the same day, especially during the initial phases of the pandemic [30]. Furthermore, lower case counts at the end of an outbreak and on weekends owing to the closure of testing centers were also considered using 7-day average metrics.

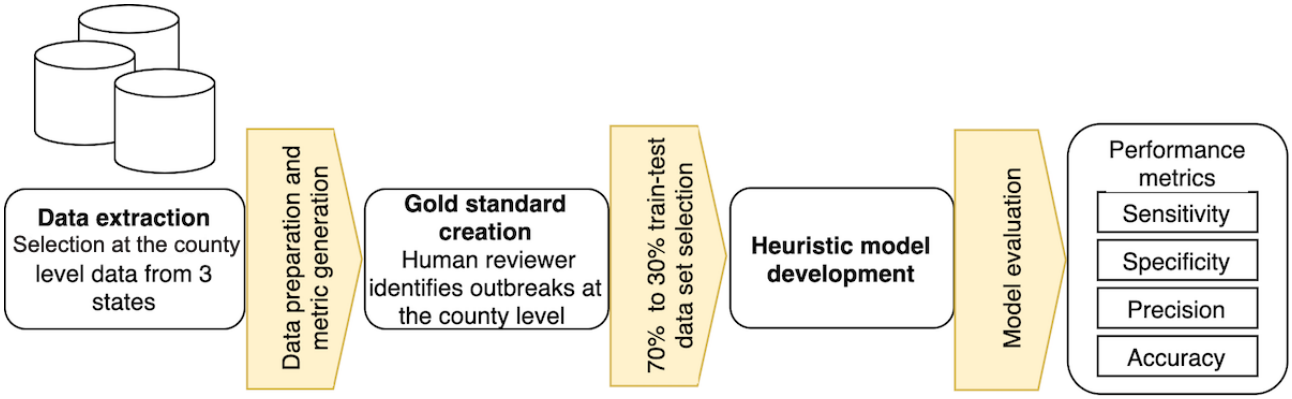
XSL•FO
RenderX

Model Building

We created a heuristic outbreak prediction model using the training data sets obtained from all 3 states and evaluated its performance across the holdout test data sets and the gold standard. For each county, data collected between March 1 and August 31, 2020, were considered the training data set, while data collected between September 1 and October 31, 2020, were considered the test data set. As a preliminary step toward model development, we considered features used in other common models, including the susceptible-infected-recovered epidemic [31] and time delay [32] models, severity of lockdown measures [33], cumulative cases (both reported and not reported) [34], and daily test reports [35]. Furthermore, predictive models for infectious diseases, such as susceptible-infected-recovered models, provide guidance on disease transmission and outbreak causation. The State of Wisconsin's COVID-19 dashboard used a Case Rate metric defined as a per capita running 7-day sum of the case counts per county per day [36]. Case Rate standardizes COVID-19 severity across counties of differing populations while also accounting for data lags and providing insight on transmission. We plotted Case Rate vs Indiana county populations to generate a general trendline that could differentiate between "outbreak detected" or "outbreak not detected" days. Our logarithmic graph semiaccurately depicted a horizontal line that separated outbreak days. The following steps were undertaken to leverage and apply the trendline results on states and counties with various populations.

We started building the model by dividing counties on the basis of population size, initially at 100,000 population intervals. Since Case Rate is more sensitive to less populated counties, we added intervals for counties with less than 100,000 people. Each population interval was allocated an assigned Case Rate baseline value that served as a binary indicator for outbreak determination. We implemented a criterion where an outbreak was underway in counties if they were 4 SDs above the cumulative case count mean to account for data lag. As depicted in the system flow diagram (Figure 1), we established these parameter values and trained the model rules with the training data sets (data reported between March 1 and August 31, 2020). The train-to-test partition was approximately 71% to 29%, respectively, which is close to optimal for large data sets [37]. Then, the model was tested against the gold standard with the test data sets (data reported between September 1 and October 31, 2020). Figure 1 shows a flow diagram depicting our study approach.

Figure 1. Flow diagram providing an overview of the study methodology.



Results

Results Overview

We collected data on a total of 249 counties from across all 3 states. Table 1 presents descriptive statistics for each state, including the number of counties, population sizes, and urbanization to highlight each state's fundamental differences [26]. Previous studies have identified multiple factors in determining urban vs rural areas, including the total population, population density, and commuting flow [38].

Indiana and Iowa have similar county population distributions, with both having a majority of counties with less than 100,000 people. However, Indiana has more midsized counties with its largest county having almost 1 million people, while California has several counties having populations of more than 1 million people. Moreover, California has the highest percentage of the urban population (94.95%), with Indiana (72.44%) and Iowa (64.02%) far behind.

Figure 2 provides an example visualization of outbreak determination in Cass County, Indiana, and Santa Barbara County, California, for gold standard preparation. Cass and Santa Barbara counties have populations of 37,689 and 446,499 people, respectively [26].

XSL•FO
RenderX

Table 1. State and county population sizes and population statistics based on census counts.

Census counts	Indiana	Iowa	California
County-level statistics			
Counties, n	92	99	58
Counties where population is <100,000 people, n	75	93	23
Counties where the population is ≥100,000 and <500,000 people, n	16	6	19
Counties where the population is ≥500,000 and <1,000,000 people, n	1	0	7
Counties where the population is >1,000,000 people, n	0	0	9
Population of the smallest county (people), n	5875	3602	1129
Population of the largest county (people), n	964,582	490,161	10,039,107
Urban population, %	72.44	64.02	94.95
Household income (US $), median	59,892	68,718	70,489
Case count per day, mean (SD)	7.98 (21.02)	6.18 (15.75)	70.33 (231.86)

Figure 2. Visualization of the COVID-19 case counts in Cass County, Indiana, and Santa Barbara County, California, between March 1 and October 31, 2020. Outbreak days are indicated in red, and normal days are indicated in black.

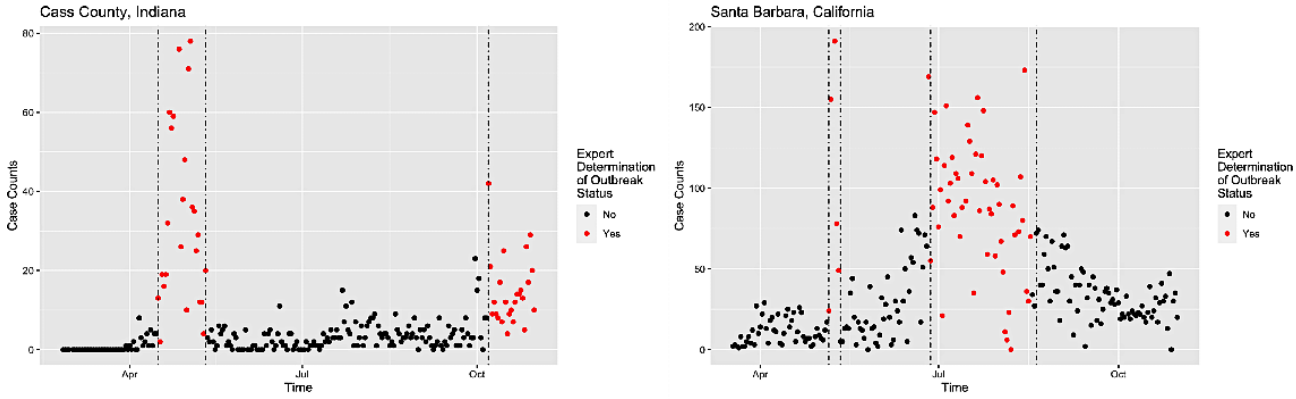


Table 2 shows the prevalence of the number of outbreaks and their durations in each state over the training, test, and total time periods. In Indiana and Iowa, the number of outbreaks doubled from the training to the test date range, despite the training data set being almost 3-fold as large as the test data set. Furthermore, the percentage of days in an outbreak between the training and test ranges quadrupled to 22.6% and 20.1% for Indiana and Iowa, respectively. The percentage of outbreak days in California remained relatively stable, while the average outbreak duration decreased from 47 days to 19 days. Because counties were our unit of analysis and since California had fewer and more populated counties than Indiana or Iowa, we believe that these factors contributed to the reduced number of outbreaks in California.

Table 2. COVID-19 outbreak prevalence descriptors from the gold standard. Indiana, Iowa, and California data sets were divided by training (March 1 to August 31, 2020), test (September 1 to October 31, 2020), and total (March 1 to October 31, 2020) date ranges to characterize the outbreak periods.

Prevalence descriptors	Indiana			Iowa			California		
	Training	Test	Total	Training	Test	Total	Training	Test	Total
Outbreaks, n	26	65	83	43	85	114	35	26	40
Outbreak duration (days), mean	25.00	19.18	22.86	18.18	14.62	15.79	47.29	19.31	53.92
Total outbreak days, n	650	1247	1897	727	1199	1926	1655	502	2157
Outbreak days, %	3.86	22.59	8.45	4.01	20.15	7.97	15.59	14.43	5.24

Model Rules

Using the aforementioned data sets, we developed model parameters to predict COVID-19 outbreaks.

Figure 3 shows a top-down decision tree for our model's behavior. Rules and the assigned case rate associated with each population band used in the decision-making process are further outlined in Multimedia Appendix 1. As shown in Figure 3, the heuristic model determined that counties experienced an outbreak through the following methods:

XSL•FO
RenderX

1. For the specified population band, a county's case rate on a given day was greater than the minimum case rate assigned to that population.
2. The county's case count on a specific day was greater than 12 and was 4 SDs above the rolling mean county-level COVID-19 case count.
3. If a county met either requirement on a specific day, that county was considered to be "in outbreak."

By combining these rules with the previously developed gold standard, a confusion matrix was utilized to analyze the model's performance.

Figure 3. Decision-making process of our heuristic model.

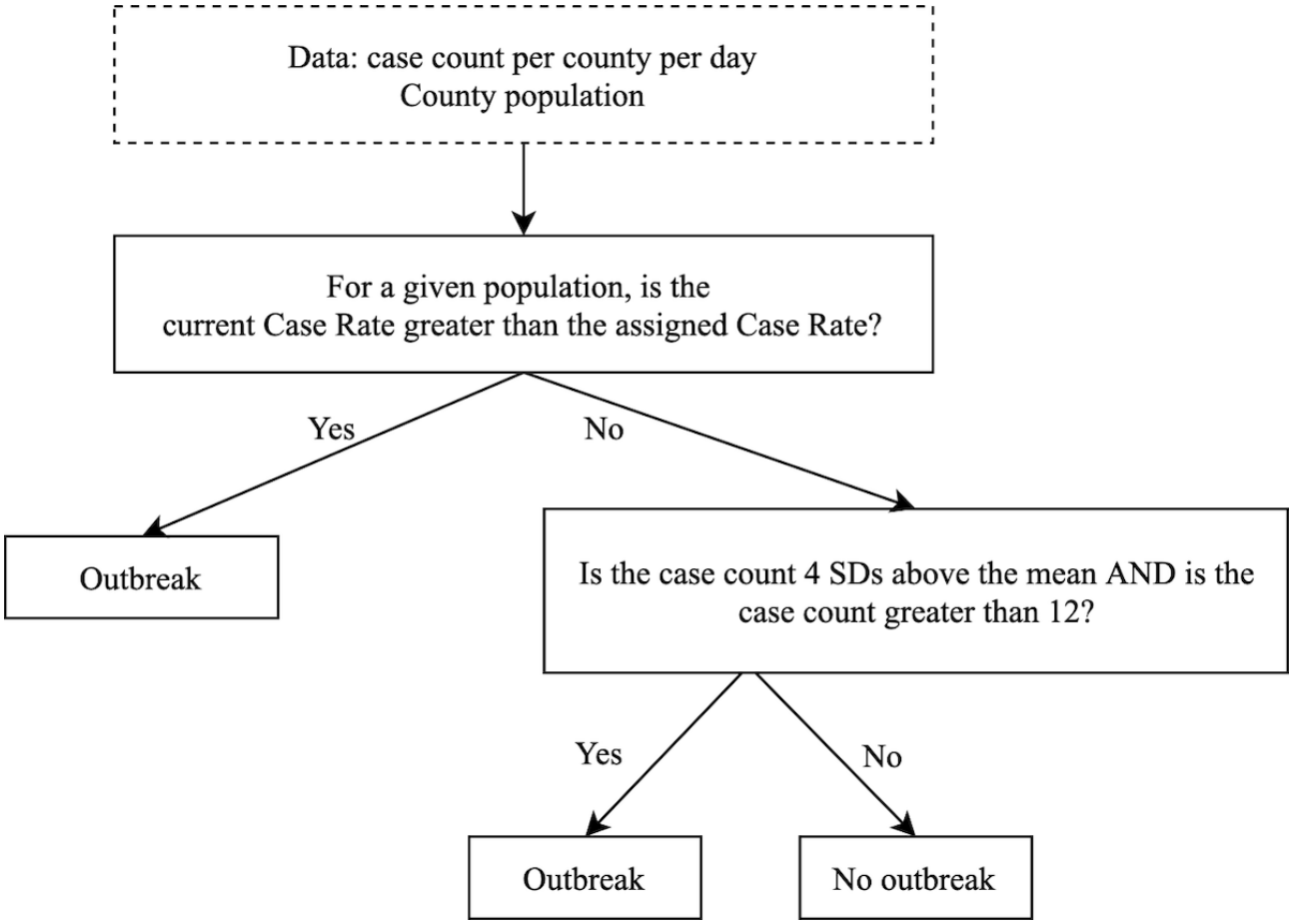


Table 3 shows the results of the confusion matrix when the prediction model was applied to the curated gold standard during the test date range from September 1 to October 31, 2020. Sensitivity is the proportion of correctly identified positives, while specificity is the proportion of correctly identified negatives. All 4 key confusion matrix statistics—sensitivity, specificity, precision, and accuracy—were above 80% in each state during the test range. Model specificity and accuracy were >94% for each state. This was attributed to most days being classified as true negatives, which are fundamentally more straightforward to detect than true positives. Model sensitivity for Indiana and Iowa was 10% greater than that of California. However, their precision was 11% and 7% lower, respectively. For Indiana and Iowa, this implies that the model computed fewer false-negative readings, which could be attributed to having increasingly prolonged outbreaks (Table 1). California's higher precision but lower sensitivity implies that the model was more precise in predicting when outbreaks occurred but less successful in capturing all outbreaks.

Table 3. Results of the model with the test data set relative to the gold standard.

Performance parameters	Indiana	Iowa	California
Test date range	September 1 to October 31, 2020	September 1 to October 31, 2020	September 1 to October 31, 2020
Sensitivity, %	92.33	90.05	80.86
Specificity, %	95.56	97.40	99.57
Precision, %	85.04	89.83	96.96
Accuracy, %	94.86	95.91	96.85

XSL•FO
RenderX

Discussion

Principal Findings

Our efforts resulted in the development of a heuristic model capable of detecting COVID-19 outbreaks, with predictive measures between 80% and 99%. The model had sensitivities of 92%, 90%, and 81% for Indiana, Iowa, and California, respectively. This indicates that the model was capable of identifying a clear majority of outbreaks across each state. The model also reported precision scores of 85%, 89%, and 96% for Indiana, Iowa, and California, respectively, which indicated that most positive predictions made by the model were accurate. These performance metrics indicate that the model is fit for use in real-life settings. Additionally, the training and test periods displayed distinct outbreak characteristics owing to the increased spread of COVID-19.

These performance metrics are also notable, considering that the prevalence of outbreaks in training data sets was considerably low and could have resulted in weak predictive models had we used more traditional classification-based modeling approaches, which significantly underperform when trained using unbalanced data sets [39,40]. As the pandemic progressed, each state attempted to enhance their data reporting systems. As described by Khakharia et al [41], some regions reported sudden and significant changes in case counts, making it difficult for models to forecast future cases. Though outbreaks are not fundamentally different, the training and test data sets can be characterized separately. Despite the test range being shorter, Indiana and Iowa both reported twice as many outbreaks during the test period. This can be attributed to the second wave of COVID-19 that occurred during the test period as schools resumed, governors relaxed state lockdown laws, and people returned to work [42]. For example, California was one of the last states to begin lifting restrictions in midsized and large counties, which may have contributed to relatively fewer outbreaks than those in Indiana and Iowa [43,44]. Thus, counties re-entered or for the first time realized outbreak periods during the test period.

California remains a state of interest owing to the characterization of its outbreaks as well as predictive performance results on the holdout test data set. Unlike Indiana and Iowa, California has several counties with populations of over 1,000,000 people; furthermore, it was the only state with fewer outbreaks and a low percentage of outbreak days between the training and test periods. The California model revealed a significantly lower sensitivity but higher precision. Thus, to Indiana and Iowa, the California model captured proportionally fewer outbreaks but predicted the subset with greater precision.

This parsimonious prediction model is easily replicable in other states, as it only utilizes county population and COVID-19 cases per day per county data. States can detect and predict outbreaks with high accuracy by following the model's rules. Current outbreak prediction approaches are based on machine learning algorithms. Though they generally have very high accuracies, these models incorporate a variety of data points and can overfit models [22]. The heuristic model's data simplicity enables it to be easily implemented in other regions, especially those with limited reported systems. It is also an understandable and accurate method to relay a county's current state of COVID-19 to the general public, who are not as informed in health metrics. In addition to public and internal communication, forecasting models can be applied to aid in outbreak preparation and community mitigation methods [45].

In addition to a high-performing heuristic model, our efforts also led to the development of a well-curated gold standard data set consisting of outbreak status for each county on a day-to-day basis. This data set is shown in Multimedia Appendix 2 to facilitate additional studies on this important issue.

Limitations

Our study was impacted by limitations in data collection systems currently deployed by each state. The inconsistency of data reporting presented a significant systematic challenge for model building activities. For instance, states closed most COVID-19 testing centers on weekends, which led to lower case count values on Saturdays and Sundays. Further, many states did not publish most of their own COVID-19 data, which led us to obtain data on cases per day per county from the New York Times instead of a state's Department of Health, the latter being more accurate. The New York Times would retroactively change case data, making it more unreliable since there were days with negative values.

The lack of previous studies on curating gold standards on disease outbreaks also presents limitations. With no industry standard on defining an outbreak, we created the gold standard on the basis of intuition and the aforementioned specified criteria. Therefore, this process could have been subject to potential confounders, which may have influenced our model's results. Furthermore, the rule-based model approach is subject to several limitations. Since the model incorporated a 7-day moving Case Rate, there was a lag at the tails of outbreaks as the increased case counts were not initially detected. Even with a parsimonious approach, the parameters derived from our results can greatly differ when applied to other regions. This uncertainty, resulting from parameters, social mandates, and vaccination, is a feature of any prediction model. We helped lessen this uncertainty through our generalizable approach demonstrated in diverse states.

Future Prospects

The ongoing COVID-19 pandemic has led most major institutions to allocate tremendous resources for its resolution. The model would benefit from a larger sample size of US states, and possibly regions worldwide, to test its generalizability on a more expansive scale. Additionally, we could expand the model's data range for the third wave of cases and as the COVID-19 vaccine is distributed to a majority of the population, to determine its functionality beyond the scope of this study. Our results could also be translated to provide a clearer epidemiological outlook of diseases. Since the model can predict outbreaks with high accuracy, it could be tested on historical COVID-19 data to determine when most outbreaks occurred easily in a particular region. Moreover, trends and patterns were found across outbreaks among various factors such as lockdown policies, air pollution levels, and civilian obedience.

XSL•FO
RenderX

Understanding the causes of outbreaks presents interesting findings related to public policy adaptation in current and future situations.

Conclusions

This study presents an accurate, generalizable, and explainable COVID-19 outbreak prediction model. The model reported sensitivity scores of >90% in Indiana and Iowa and >80% in California. Furthermore, model specificity and accuracy scores were >94% in every state. These results, coupled with the minimal data inputs required, creates an explainable and easy-to-implement model that governments and policymakers can utilize to assess COVID-19 severity across diverse geographic regions. Future studies are required to test the model in other states and countries by using more recent data. Moreover, the model should be used to identify outbreaks to investigate correlations among external factors such as socioeconomic risks, air pollution, county-level laws, and outbreak development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Population bands with their respective minimum case rates.
[XLS File (Microsoft Excel File), 26 KB - jmir_v23i7e28812_app1.xls]

Multimedia Appendix 2
Outbreak gold standard for each county in California, Indiana, and Iowa per day.
[XLSX File (Microsoft Excel File), 38 KB - jmir_v23i7e28812_app2.xlsx]

References

1. Estiri H, Strasser ZH, Klann JG, Naseri P, Wagholikar KB, Murphy SN. Predicting COVID-19 mortality with electronic medical records. NPJ Digit Med 2021 Feb 04;4(1):15 [FREE Full text] [doi: 10.1038/s41746-021-00383-x] [Medline: 33542473]
2. Hartley DM, Perencevich EN. Public Health Interventions for COVID-19: Emerging Evidence and Implications for an Evolving Public Health Crisis. JAMA 2020 May 19;323(19):1908-1909. [doi: 10.1001/jama.2020.5910] [Medline: 32275299]
3. Baker S, Bloom N, Davis S, Terry S. COVID-Induced Economic Uncertainty. National Bureau of Economic Research. 2020. URL: https://www.nber.org/system/files/working_papers/w26983/w26983.pdf [accessed 2021-07-13]
4. McNeil D. Coronavirus Has Become a Pandemic. The New York Times. URL: https://www.nytimes.com/2020/03/11/health/coronavirus-pandemic-who.html [accessed 2021-07-13]
5. Coronavirus (Covid-19) Data. The New York Times Developer Network. URL: https://developer.nytimes.com/covid [accessed 2021-07-13]
6. Liu Q, Luo D, Haase JE, Guo Q, Wang XQ, Liu S, et al. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. Lancet Glob Health 2020 Jun;8(6):e790-e798 [FREE Full text] [doi: 10.1016/S2214-109X(20)30204-7] [Medline: 32573443]
7. Cutler DM, Summers LH. The COVID-19 Pandemic and the $16 Trillion Virus. JAMA 2020 Oct 20;324(15):1495-1496 [FREE Full text] [doi: 10.1001/jama.2020.19759] [Medline: 33044484]
8. Rubin R. NIH Launches Platform to Serve as Depository for COVID-19 Medical Data. JAMA 2020 Jul 28;324(4):326. [doi: 10.1001/jama.2020.12646] [Medline: 32720992]
9. Bennett T, Moffitt R, Hajagos J, Amor B, Anand A, Bissell MM, et al. The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction. medRxiv. Preprint posted online January 23, 2021 [FREE Full text] [doi: 10.1101/2021.01.12.21249511] [Medline: 33469592]
10. Dixon B, Grannis S, McAndrews C, Broyles AA, Mikels-Carrasco W, Wiensch A, et al. Leveraging Data Visualization and a Statewide Health Information Exchange to Support COVID-19 Surveillance and Response: Application of Public Health Informatics. J Am Med Inform Assoc 2021 Jan 22:ocab004 [FREE Full text] [doi: 10.1093/jamia/ocab004] [Medline: 33480419]
11. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
12. Jewell NP, Lewnard JA, Jewell BL. Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic. Ann Intern Med 2020 Aug 04;173(3):226-227 [FREE Full text] [doi: 10.7326/M20-1565] [Medline: 32289150]
13. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]

XSL•FO
RenderX

14. Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, et al. BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. N Engl J Med 2021 Apr 15;384(15):1412-1423 [FREE Full text] [doi: 10.1056/NEJMoa2101765] [Medline: 33626250]
15. Del Rio C, Malani P. COVID-19 in 2021-Continuing Uncertainty. JAMA 2021 Apr 13;325(14):1389-1390. [doi: 10.1001/jama.2021.3760] [Medline: 33662092]
16. COVID Data Tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#datatracker-home [accessed 2021-07-13]
17. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. Ann Intern Med 2020 Jul 07;173(1):21-28 [FREE Full text] [doi: 10.7326/M20-1260] [Medline: 32259197]
18. Dai Y, Wang J. Identifying the outbreak signal of COVID-19 before the response of the traditional disease monitoring system. PLoS Negl Trop Dis 2020 Oct;14(10):e0008758 [FREE Full text] [doi: 10.1371/journal.pntd.0008758] [Medline: 33001985]
19. Adnan M, Gao X, Bai X, Newbern E, Sherwood J, Jones N, et al. Potential Early Identification of a Large Campylobacter Outbreak Using Alternative Surveillance Data Sources: Autoregressive Modelling and Spatiotemporal Clustering. JMIR Public Health Surveill 2020 Sep 17;6(3):e18281 [FREE Full text] [doi: 10.2196/18281] [Medline: 32940617]
20. Tseng Y, Shih Y. Developing epidemic forecasting models to assist disease surveillance for influenza with electronic health records. Int J Comput Appl 2019 Aug 05;42(6):616-621. [doi: 10.1080/1206212x.2019.1633762]
21. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google Search Trends Predicting Disease Outbreaks: An Analysis from India. Healthc Inform Res 2018 Oct;24(4):300-308 [FREE Full text] [doi: 10.4258/hir.2018.24.4.300] [Medline: 30443418]
22. Dexter G, Grannis S, Dixon B, Kasthurirathne S. Generalization of Machine Learning Approaches to Identify Notifiable Conditions from a Statewide Health Information Exchange. AMIA Jt Summits Transl Sci Proc 2020;2020:152-161 [FREE Full text] [Medline: 32477634]
23. Sáez C, Romero N, Conejero J, García-Gómez JM. Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset. J Am Med Inform Assoc 2021 Feb 15;28(2):360-364 [FREE Full text] [doi: 10.1093/jamia/ocaa258] [Medline: 33027509]
24. Ghorbani A, Abid A, Zou J. Interpretation of Neural Networks Is Fragile. AAAI 2019 Jul 17;33:3681-3688. [doi: 10.1609/aaai.v33i01.33013681]
25. Vandekerckhove J, Matzke D, Wagenmakers E. Model comparisonthe principle of parsimony. In: The Oxford handbook of computationalmathematical psychology. New York, NY: Oxford University Press; 2015:300-319.
26. County Population Totals: 2010-2019. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html [accessed 2021-07-13]
27. State Reporting Assessments. The COVID Tracking Project. URL: https://covidtracking.com/about-data/state-reporting-assessments [accessed 2021-07-13]
28. MPH helps Indiana fight COVID-19 with data. State of Indiana. URL: https://www.in.gov/mph/1194.htm [accessed 2021-07-13]
29. Ripley BD. The R Project in Statistical Computing. MSOR Connections 2001 Feb;1(1):23-25. [doi: 10.11120/msor.2001.01010023]
30. Maxmen A. Why the United States is having a coronavirus data crisis. Nature 2020 Sep;585(7823):13-14. [doi: 10.1038/d41586-020-02478-z] [Medline: 32843755]
31. Yoshida N, Hara T. Global stability of a delayed SIR epidemic model with density dependent birth and death rates. J Comput Appl Math 2007 Apr 23;201(2):339-347 [FREE Full text] [doi: 10.1016/j.cam.2005.12.034]
32. DrugTherapeutics Bulletin. An introduction to patient decision aids. BMJ 2013 Jul 23;347:f4147 [FREE Full text] [Medline: 23881944]
33. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
34. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 2020 Apr;93:211-216 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.058] [Medline: 32145465]
35. Casella F. Can the COVID-19 epidemic be controlled on the basis of daily test reports? 2020. Astrophysics Data System. 2020. URL: https://ui.adsabs.harvard.edu/abs/2020arXiv200306967C [accessed 2020-03-01]
36. COVID-19: Activity Level by Region and County. Wisconsin Department of Health Services. URL: https://www.dhs.wisconsin.gov/covid-19/disease.htm [accessed 2021-07-13]
37. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics 2011 Apr 08;4:31 [FREE Full text] [doi: 10.1186/1755-8794-4-31] [Medline: 21477282]
38. Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in U.S. epidemiologic studies. J Urban Health 2006 Mar;83(2):162-175 [FREE Full text] [doi: 10.1007/s11524-005-9016-3] [Medline: 16736366]

XSL•FO
RenderX

39. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 2016 Apr 22;5(4):221-232. [doi: 10.1007/s13748-016-0094-0]
40. Kasthurirathne SN, Grannis SJ. Machine Learning Approaches to Identify Nicknames from A Statewide Health Information Exchange. AMIA Jt Summits Transl Sci Proc 2019;2019:639-647 [FREE Full text] [Medline: 31259019]
41. Khakharia A, Shah V, Jain S, Shah J, Tiwari A, Daphal P, et al. Outbreak Prediction of COVID-19 Patients for Dense and Populated Countries Using Machine Learning. SSRN Journal 2020. [doi: 10.2139/ssrn.3655759]
42. Maragakis L. Coronavirus Second Wave? Why Cases Increase. Johns Hopkins Medicine. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus [accessed 2021-07-13]
43. Each State's COVID-19 Reopening and Reclosing Plans and Mask Requirements. National Academy for State Health Policy. URL: https://www.nashp.org/governors-prioritize-health-for-all/ [accessed 2021-02-22]
44. Li Y, Zhang R, Zhao J, Molina MJ. Understanding transmission and intervention for the COVID-19 pandemic in the United States. Sci Total Environ 2020 Dec 15;748:141560 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.141560] [Medline: 32798884]
45. Lutz CS, Huynh MP, Schroeder M, Anyatonwu S, Dahlgren FS, Danyluk G, et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. BMC Public Health 2019 Dec 10;19(1):1659 [FREE Full text] [doi: 10.1186/s12889-019-7966-8] [Medline: 31823751]

Edited by C Basch; submitted 23.03.21; peer-reviewed by R Foraker, S Soshnikov, S Pesälä; comments to author 05.05.21; revised version received 02.06.21; accepted 12.06.21; published 26.07.21.

Please cite as:
Gupta AK, Grannis SJ, Kasthurirathne SN
Evaluation of a Parsimonious COVID-19 Outbreak Prediction Model: Heuristic Modeling Approach Using Publicly Available Data Sets
J Med Internet Res 2021;23(7):e28812
URL: https://www.jmir.org/2021/7/e28812
doi:10.2196/28812
PMID:34156964

©Agrayan K Gupta, Shaun J Grannis, Suranga N Kasthurirathne. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Awareness and Utilization of 937-Telephone Health Services in the Kingdom of Saudi Arabia Before and During the COVID-19 Pandemic: Longitudinal Study

Saja A Al-Rayes[1], PhD; Arwa Alumran[1], PhD; Duaa Aljabri[1], PhD; Afnan Aljaffary[1], PhD; Ethar Aldoukhi[1], BSc; Zainab Alahmedalyousif[1], BSc; Reem Al Madani[2], MSc

[1]Health Information Management and Technology Department, College of Public Health, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia

[2]Risk Management Unit, Directorate of Quality and Safety, King Fahah Hospital of the University, Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia

Corresponding Author:
Afnan Aljaffary, PhD
Health Information Management and Technology Department
College of Public Health
Imam Abdulrahman bin Faisal University
King Faisal Road
Dammam
Saudi Arabia
Phone: 966 1333333766
Email: aaljaffari@iau.edu.sa

Abstract

Background: Telehealth plays a key role in supporting health care systems and influencing methods of health care delivery. Government laws and medical operating protocols have been largely modified to provide remote care to reduce social contact and ensure a safer patient environment. In the Kingdom of Saudi Arabia (KSA), the Ministry of Health (MOH) introduced several forms of telemedicine as alternatives to face-to-face consultations in clinical settings.

Objective: This study aimed to assess the awareness and utilization of telehealth services before and during the COVID-19 outbreak in the KSA.

Methods: In this longitudinal study, we compared the awareness and utilization of 937-telephone health services (ie, a toll-free telephone service to provide medical and administrative health care services at any time for the population) before and during the COVID-19 outbreak in the KSA. Using a convenience sampling technique, a validated web-based questionnaire was distributed on social media platforms (Facebook, Twitter, and WhatsApp) at 2 timepoints: before (February 2019) and during (May 2020) the COVID-19 pandemic.

Results: The study sample comprised a total of 1961 participants who completed the questionnaire before (n=1303, 66%) and during (n=658, 33%) the COVID-19 pandemic. Both awareness (before=46% vs during=78%) and utilization (before=42% vs during=48%) of the 937-telephone health services increased significantly during the pandemic (P<.001). No significant association of the awareness or utilization of 937-telephone health services before and during the COVID-19 pandemic was found with respect to the participants' age, education level, having children, or having any chronic disease.

Conclusions: Our findings indicate significant increases in the awareness and utilization of 937-telephone health services during the early days of the COVID-19 pandemic, suggesting an increase in public acceptance of the service and providing evidence of an equitable telemedicine service for the population. Further studies are needed to provide a deeper understanding of the barriers and facilitators to the use of 937-telephone health services for different groups of the population.

(J Med Internet Res 2021;23(7):e27618) doi:10.2196/27618

KEYWORDS

awareness; COVID-19; Kingdom of Saudi Arabia; telehealth; telemedicine; utilization

XSL•FO
RenderX

Introduction

The COVID-19 pandemic has transformed the manner in which health care services are provided to patients worldwide. With the rapid spread of COVID-19 from 1 city (Wuhan, China) to almost all countries, governments worldwide implemented restrictive measures to reduce the contamination risk. Many countries adopted extreme prevention methods to prevent disease transmission, such as self-isolation and quarantine, social distancing, prioritizing health centers for COVID-19 cases, closing borders, and the complete lockdown of cities. In an attempt to reduce the opportunity of disease transmission, medical operating protocols have been largely modified to comply with control measures. Thus, the delivery of health care services has shifted toward the use of telemedicine and remote services for medical consultations [1-5].

Telemedicine was first described in the 1970s and was defined as the process of providing health care from a distance through the use of communications technology [6,7]. Telephone consultations can be considered a rudimentary form of telemedicine [8]. Telemedicine can be described as "the process whereby patients receive medical advice by one or more qualified healthcare professionals via the telephone" [9].

In the Kingdom of Saudi Arabia (KSA), the Ministry of Health (MOH) introduced several forms of telemedicine as alternatives to face-to-face consultations in clinical settings. More than 19 mobile apps have been released to provide web-based health services to smartphone users, who constitute 96% of the population [10]. Another key initiative to improve health care access for those without an internet connection was the establishment of 937—a free, confidential telephone service that provides medical and administrative health care services at any time. These 937-telephone health services are available 24/7 throughout the year and are provided in both Arabic and English. Calls are answered by health care professionals who can ask questions to assess the health problem and provide the caller advice on whether to handle the problem at home or to go to a hospital emergency room. The service provides four main subservices: (1) medical consultation, (2) reporting complaints about public and private health facilities, (3) appointment requests for primary health centers and smoking cessation clinics, and (4) responding to inquiries such as requesting information about infectious diseases, toxins, and medicines and providing technical support for MOH mobile apps.

The 937-telephone health services are not limited to the telephone number "937" but are rather linked with a mobile app with an instant messaging feature, which allows patients to send photographs or videos during medical consultations (eg, of a skin rash). The service is also accessible for deaf people and those with hearing impairment through a special mobile app "Ishara" (ie, "sign"), through which a service provider launches a video call to connect the patient to the 937-telephone health care services and provides a sign-language translation.

In the months before the pandemic, the 937-telephone health services were not in effective use in the KSA, with only approximately 46% of the population reporting awareness of the service and only 20% using it [11]. Given that the first part of this study was conducted before the COVID-19 outbreak, we were able to assess whether awareness and utilization of the 937-telephone service had increased during the pandemic. This study aimed to assess the awareness and utilization of the 937-telephone health services before and during the COVID-19 outbreak in the KSA, and we hypothesized that awareness and utilization of the 937-telephone health services increased during the pandemic.

Methods

Study Design

This study used a longitudinal approach, where data were collected at two timepoints: before (February 2019) and during the COVID-19 pandemic (May 2020).

Participants

The survey included 2 cohorts of citizens and residents of the KSA. Participants were selected using a nonprobability sampling technique, specifically convenience sampling.

Variables

The questionnaire measured awareness and utilization of the 937-telephone health services in the KSA. Awareness was measured by asking, "Are you aware of the Ministry of Health 937-telephone health service (yes vs no)?" Utilization was measured by asking, "Have you ever contacted 937-telephone health services (yes vs no)?" Participants who answered "yes" to the second question were asked a follow-up question regarding the purpose of their call (eg, for a medical consultation, complaints, appointments, or a general inquiry).

In addition, data on independent variables were collected to assess their association with awareness and utilization. These included data on age (age groups: ≤19 years, 20-29 years, 30-39 years, 40-49 years, 50-59 years, and ≥60 years), gender (male or female), level of education (at least high school, diploma, bachelor's degree, or graduate degree), area of residence (Eastern, Western, Northern, Southern, and Central regions), nationality (Saudi citizen or resident), having children younger than 18 years old (yes vs no), and previous diagnoses of chronic disease (yes vs no).

Data Source or Instrument

As digital communication channels have been increasingly used to recruit study participants in recent studies, especially during the COVID-19 pandemic, we used a web-based questionnaire for data collection. Three methods were used to validate the survey: face validity (to test the clarity and relevancy of the content and the scale overall layout), content validity, and linguistic validity. The validation process has been previously described by Alrayes et al [11].

Sample Size

The questionnaire was distributed on social media platforms (ie, Facebook, Twitter, and WhatsApp) with no inclusion or exclusion criteria; hence, formal sample size calculations were not feasible. As an alternative, we followed a previously

XSL•FO
RenderX

recommended guideline [12] for sample sizes: 100="fair," 200="good," 500="very good," and >1000="excellent."

Statistical Analysis

The data were analyzed using SPSS (IBM Corp) [13]. Demographic characteristics, awareness, and utilization were analyzed using descriptive analysis. Chi-square tests were used to assess the difference in awareness and utilization of the 937-telephone health services before and during the pandemic. In addition, factors influencing the awareness and utilization of the 937-telephone health services were assessed using chi-square tests, and *P* values less than .05 were considered significant for all analyses.

Ethical Considerations

All participants were asked for their voluntary participation and consent before filling out the questionnaire and were free to terminate the survey at any time. Ethics approval was obtained from the institutional review board of Imam Abdulrahman Bin Faisal University (protocol# IRB-UGS-2018-03-294).

Results

Participants

The study sample included 1961 participants who completed the questionnaire before (n=1303, 66%) and during (n=658, 34%) the COVID-19 pandemic. Nonsignificant differences were observed between the population cohorts, which indicated that the study samples were homogenous.

Descriptive Data

Most study participants were women (n=1488, 76%), and approximately one-third were 20-29 years old (n=628, 32%). Half of the participants indicated that they had at least 1 child (n=1008, 51%), and most did not have a chronic disease (n=1705, 87%). The majority of study participants resided in the Eastern region (n=1624, 83%) (Table 1).

XSL•FO
RenderX

Table 1. Awareness of 937-telephone health services in the Kingdom of Saudi Arabia before and during COVID-19 pandemic.

Variables	Overall (n=1961), n (%)	Aware (n=1113), n (%)	Awareness of the 937-telephone health service (n=1113)		Chi-square (*df*)	*P* value
			Before the COVID-19 pandemic (n=601), n (%)	During the COVID-19 pandemic (n=512), n (%)		
Gender					11.5 (1)	.001[a]
Female	1488 (76)	813 (73)	464 (77)	349 (68)		
Male	473 (24)	300 (27)	137 (23)	163 (32)		
Age (years)					5.2 (5)	.39
≤19	117 (6)	40 (4)	18 (3)	22 (4)		
20-29	628 (32)	332 (30)	187 (31)	145 (28)		
30-39	454 (23)	300 (27)	168 (28)	132 (26)		
40-49	443 (23)	258 (23)	136 (23)	122 (24)		
50-59	262 (13)	153 (14)	80 (13)	73 (14)		
≥60	57 (3)	30 (3)	12 (2)	18 (4)		
Having children						
Yes	1008 (51)	632 (57)	352 (59)	280 (55)	1.7 (1)	.19
No	953 (49)	481 (43)	249 (41)	232 (45)		
Nationality						
Saudi citizen	1907 (97)	1090 (98)	594 (99)	496 (97)	5.3 (1)	.02[a]
Resident	54 (3)	23 (2)	7 (1)	16 (3)		
Area of residence						
East	1624 (83)	916 (82)	462 (77)	454 (89)	45.4 (3)	<.001[a]
Central	170 (9)	113 (10)	94 (16)	19 (4)		
West	112 (5)	54 (5)	32 (5)	22 (4)		
North and south	55 (3)	30 (3)	13 (2)	17 (3)		
Education level					3.4 (3)	.33
High school or lower	399 (20)	199 (18)	97 (16)	102 (20)		
Diploma	225 (12)	136 (12)	75 (13)	61 (12)		
Bachelor's degree	1,154 (59)	641 (58)	349 (58)	292 (57)		
Graduate degree	183 (9)	137 (12)	80 (13)	57 (11)		
Diagnosis of a chronic disease					1.0 (1)	.33
Yes	256 (13)	153 (14)	77 (13)	76 (15)		
No	1705 (87)	960 (86)	524 (87)	436 (85)		

[a]Significant differences.

Awareness and Utilization of the 937-Telephone Service

Among the participants who responded during the pandemic, 512 (78%) indicated that they were aware of the 937-telephone health services, while only 46% of those who responded before the pandemic were aware of the service, which indicates a significant increase in awareness since the pandemic emerged in the KSA (χ^2_1=178.9; *P*<.001) (Table 2). Awareness of the 937-telephone health services was higher among women (68%), those aged 20-29 years (28%), and those with at least 1 child (55%) (Table 1).

XSL•FO
RenderX

Table 2. Influence of the COVID-19 pandemic on the awareness and utilization of the 937-telephone health services in the Kingdom of Saudi Arabia.

Variables	Overall, n (%)	Before the COVID-19 pandemic, n (%)	During the COVID-19 pandemic, n (%)	Chi-square (*df*)	*P* value
Aware of the 937-telephone health services				178.9 (1)	<.001[a]
Yes	1113 (57)	601 (46)	512 (78)		
No	848 (43)	702 (53)	146 (22)		
Using the 937-telephone health services				76.6 (1)	<.001[a]
Yes	509 (26)	258 (42)	251(48)		
No	1452 (74)	1045 (58)	407 (52)		
Reasons for using the 937-telephone health services					
Consultations	430 (84)	218 (84)	212 (84)	0.00 (1)	.99
Complaints	307 (60)	123 (48)	184 (73)	34.9 (1)	<.001[a]
Appointments	234 (46)	92 (36)	142 (57)	22.4 (1)	<.001[a]
Inquiries	305 (60)	137 (53)	168 (67)	10.1 (1)	<.001[a]

[a]Significant differences.

Although the percentage of respondents utilizing the 937-telephone health services was similar before and during the pandemic (42% and 48%, respectively), this difference was still significant (χ^2_1=76.6; *P*<.001). Utilization during the pandemic was higher among women (n=146, 58%), those aged 30-39 years (n=85, 34%), and those with at least 1 child (n=161, 64%) (Table 3).

XSL•FO
RenderX

Table 3. Utilization of 937-telephone health services in the Kingdom of Saudi Arabia before and after COVID-19 pandemic.

Variables	Overall utilization (n=509), n (%)	Utilization of the 937-telephone health service		Chi-square (*df*)	*P* value
		Before the COVID-19 pandemic (n=258), n (%)	During the COVID-19 pandemic, (n=251), n (%)		
Gender				30.3 (1)	<.001[a]
Female	354 (70)	208 (81)	146 (58)		
Male	155 (30)	50 (19)	105 (42)		
Age (years)				4.4 (5)	.49
≤19	12 (2)	7 (3)	5 (2)		
20-29	125 (25)	68 (26)	57 (23)		
30-39	178 (35)	93 (36)	85 (34)		
40-49	119 (23)	59 (23)	60 (24)		
50-59	65 (13)	28 (11)	37 (15)		
≥60	10 (2)	3 (1)	7 (3)		
Having children				0.4 (1)	.55
Yes	333 (65)	172 (67)	161 (64)		
No	176 (35)	86 (33)	90 (36)		
Nationality				1.9 (1)	.17
Saudi citizen	504 (99)	257 (99)	247 (98)		
Resident	5 (1)	1 (1)	4 (2)		
Area of residence				26.6 (3)	<.001[a]
East	416 (82)	196 (76)	220 (88)		
Central	63 (12)	50 (19)	13 (5)		
West	20 (4)	10 (4)	10 (4)		
North and south	10 (2)	2 (1)	8 (3)		
Education level				0.9 (3)	.83
High school or lower	90 (18)	47 (18)	43 (17)		
Diploma	64 (13)	31 (12)	33 (13)		
Bachelor's degree	288 (57)	143 (55)	145 (58)		
Graduate degree	67 (13)	37 (14)	30 (12)		
Diagnosis of a chronic disease				0.3 (1)	.61
Yes	73 (14)	35 (14)	38 (15)		
No	436 (86)	223 (86)	213 (86)		

[a]Significant differences.

Reasons for Using the 937-Telephone Health Service

When assessing the reasons for using the 937-telephone health services, 4 service areas were considered: consultations, complaints, appointments, and inquiries (Table 2). Use of the service for complaints (χ^2_1=35.7; *P*<.001), appointments (χ^2_1=20.7; *P*<.001), and inquiry (χ^2_1=38.6; *P*<.001) was significantly different before and during the pandemic, while there was no significant difference in the use of the service for consultations before and during the pandemic (χ^2_1=3.4; *P*=.07).

As shown in Table 2, use of the service for complaints increased from 45% before to 63% during the COVID-19 pandemic. In addition, use of the service to schedule appointments increased from 27% before to 39% during the COVID-19 pandemic. Finally, participants' utilization of the service to make inquiries increased from 48% before to 67% during the COVID-19 pandemic.

Factors Influencing the Awareness or Utilization of the 937-Telephone Health Service

In bivariate analysis, several variables significantly influenced participants' awareness before and during the COVID-19 pandemic. These variables were the following: gender,

XSL•FO
RenderX

nationality, and area of residence. Women were generally more aware about the 937-telephone health services than men, regardless of the time point in the study (χ^2_1=11.5; *P*=.001).

Area of residence was significantly associated with awareness of the 937-telephone health services (χ^2_3=45.4; *P*<.001). There was a notable increase in awareness across all areas in the KSA after the COVID-19 pandemic emerged (Table 1). Awareness in the Eastern region increased from 77% before to 89% during the COVID-19 pandemic.

Two factors were significantly associated with the utilization of the 937-telephone health services in this study: gender (χ^2_1=30.3; *P*<.001) and area of residence (χ^2_3=26.6; *P*<.001). Female participants utilized the service more than male participants before and during the COVID-19 pandemic; however, the difference was more significant before the pandemic (81% vs 19% and 58% vs 42%, respectively). Utilization of the service increased among men during the COVID-19 pandemic. In addition, utilization of the 937-telephone health services increased significantly in the Eastern region from 76% before to 88% during the COVID-19 pandemic (Table 3).

No other variables assessed in the study showed a significant association with the awareness or utilization of the 937-telephone health services before or during the COVID-19 pandemic. These variables include age, having children, level of education, and previous diagnoses of chronic disease. This suggests relatively equal awareness and utilization of the service across the population.

Discussion

Principal Findings

The 937-telephone health service has thrived during the COVID-19 pandemic owing to its ability to deliver remote health care services to all patients in the community. Public administrations in the United States, United Kingdom, and Australia are investing in telemedicine to reduce the volume of patients visiting emergency departments for nonurgent concerns and to limit virus transmission [14]. This study aimed to compare the awareness and utilization of the 937-telephone health services before and during the COVID-19 pandemic in the KSA.

Since the COVID-19 outbreak in the KSA, the government has employed a number of policies to implement some level of lockdown, including the cancellation of regular outpatient visits, the prioritization of hospital usage for COVID-19 cases, and the expansion and improvement of available telehealth services. Several awareness programs were initiated by the MOH for public education and COVID-19 prevention, aiming to distribute proper information and increase future preparedness. One such MOH initiative was a public awareness campaign encouraging the public to use the existing 937-telephone health services for any inquiry regarding COVID-19, such as symptoms, methods of prevention, and the required course of action on the occurrence of any of its symptoms. Our study findings showed a significant increase in both awareness and utilization levels before and during the pandemic (46% vs 78% awareness and 42% vs 48% utilization). This might have resulted from the increased awareness campaigns by the government to encourage use the 937-telephone health service as an alternative approach for nonemergency concerns.

A parallel increase in the levels of awareness of the role of telehealth in relation to COVID-19 was noted in Australia [3]. Furthermore, 59% of surveyed individuals reported being aware of the 111-telephone health services of the National Health Service in the United Kingdom and 9% reported having used them [15]. Public perception of telephone consultations has been generally positive [16]. Many people value the opportunity to obtain expert advice over the telephone as an initial response to the development of new symptoms. Cited benefits include reduced waiting and travel times, convenience, and flexibility [17,18].

Our findings show that the population engaged with the 937-telephone health services differently during the pandemic. The telephone service was mostly used for consultations (84%), followed by complaints (73%), inquiries (67%), and appointment requests (57%). While utilization of the service for consultations was the same before and during the COVID-19 pandemic, its utilization for other reasons increased significantly during the pandemic. This was expected, as during the early days of the pandemic, the population might have been uncertain, owing to inconsistent, unverified, and conflicting information from various sources [19]. Thus, the ability to obtain information from an accurate source such as the MOH telephone service might have helped overcome public uncertainty.

We found that gender played a role in the awareness of 937-telephone health services, with women being more aware than men. This finding aligns with that of another study that reported that women are more likely to engage in eHealth than men [20]. In contrast, in 2017, Khatun et al [21] found that men were more aware of existing health care services, and Jung et al [22] did not find significant differences in telemedicine awareness based on gender.

Another interesting finding, although not significant, was that awareness and utilization of the 937-telephone health services were lower among older adults. Although using the 937-telephone health services does not require an internet connection or computer skills, lesser access was noted among older people. This finding is concurrent with that of Knowles et al [15] who reported that older people were less likely to use telephone-accessed health care (than younger people) and preferred face-to-face contact. Jung et al [22] reported that 77% of younger participants (aged <50 years) preferred to use telemedicine compared to 62.5% of older participants (those aged ≥50 years). Furthermore, El-Mahalli et al [23] reported that those between 30 and 50 years of age were most likely to use telemedicine. Cultural factors might have contributed to this finding as younger adults often take care of seniors, making it unnecessary for them to use the service [23]. In addition, as older people are slower in adopting new technologies than younger adults, the web-based questionnaire might not have reached the intended older population, thus causing a nonresponse bias [24]. However, as the population ages and the

XSL•FO
RenderX

prevalence of long-term conditions increases, telehealth is a particularly valuable tool that could help elderly people maintain their health and benefit those with limited mobility [25]. Thus, understanding barriers and facilitators for telehealth use among older individuals may be important.

Another interesting finding was that among study participants, area of residence was associated with the awareness of the 937-telephone health services. Those residing in the Eastern region of the KSA knew more about the service than those residing in other regions. Jung et al [22] reported that awareness of telehealth services is affected by where a person lives; for example, large cities compared to rural areas. This is most likely due to financial support in large cities [3]. In contrast, another study reported that remote regions were more aware of telehealth services [22].

Our study confirms the increase in awareness and utilization of telehealth, driven by the emergence of the COVID-19 pandemic, which has transformed health care delivery. To our knowledge, this study is the first in the KSA to assess awareness and utilization of 937-telephone health services during the COVID-19 pandemic and to compare current levels of awareness and utilization with those before the pandemic. In addition, the sample size included in the study is large (n=1961), which provides wide representation of the population.

Implications for Telephone Health Care Services During Pandemics

With the increased demand for emergency care services and government enforcement of extreme physical distancing measures, health care providers are seeking alternative strategies for providing cost-effective and safe care. Policymakers are implementing telephone consultations as a substitution for face-to-face consultations, with a goal of maintaining continuity of care in a less expensive and safer environment. Currently, the telephone service provides nonurgent care; however, the onset of the COVID-19 pandemic has accelerated the telehealth revolution and rendered the consideration of innovative models of care essential. Reliable and valid screening instruments for telephone use are required to allow health care providers to make comprehensive assessments during the remote visit. Hence, the use of web-based apps to expand telehealth use and build its capacity within the health care delivery system is important.

Limitations

However, there are several limitations of note. First, our study population consisted predominantly of women (76%), which may have biased our results; in particular, those related to differences in awareness by gender. However, studies have shown that women tend to be more proactive regarding their health than men [20]. Second, ethnicity might have a potential role in influencing the utilization of telephone services; however, because ethnicity categorization is not used in the official censuses in the KSA and thus might be culturally unacceptable, it was not be feasible to examine this issue in this study. Instead, a more culturally acceptable alternative, nationality, was considered when constructing the study questionnaire as an indirect reflection of a participant's ethnicity. In addition, area of residence is also captured and is considered a cultural categorization in the KSA. Third, the use of the convenience sampling technique, which yielded a sample in which 83% of the participants lived in the Eastern region, likely contributed to the oversampling of participants in this region, thus limiting our inference on differences observed by area of residence and awareness of the 937-telephone health services. Fourth, data on utilization and awareness during the COVID-19 pandemic were collected in March 2020. The first local COVID-19 case was reported on March 2, 2020, and shortly thereafter, the World Health Organization declared COVID-19 a pandemic (March 11, 2020) [26]. It is possible that awareness and utilization of 937-telephone health services increased over the following months. Finally, it was difficult to calculate the response rate as the questionnaire was distributed on the internet, and no power calculation for the sample size estimation was undertaken. However, the sample size fell within the normative ranges deemed "excellent" (>1000) for common research designs [12,27].

Conclusions

This study found significant increases in the awareness and utilization of 937-telephone health services in the KSA, with nonsignificant differences across population groups before and during the COVID-19 pandemic. This suggests an increase in public acceptance of the service and offers evidence for the availability of an equitable telephone-accessed health service for the population. In addition, the utilization of 937-telephone health services for medical appointments and inquiries increased significantly during rather than before the COVID-19 pandemic. Further studies are needed to provide a deeper understanding of the barriers and facilitators for the utilization of the 937-telephone health services among different population groups.

Conflicts of Interest

None declared.

References

1. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]

XSL•FO
RenderX

2. Khuroo M, Khuroo M, Khuroo M, Sofi AA, Khuroo NS. COVID-19 Vaccines: A Race Against Time in the Middle of Death and Devastation!. J Clin Exp Hepatol 2020;10(6):610-621 [FREE Full text] [doi: 10.1016/j.jceh.2020.06.003] [Medline: 32837093]
3. Fisk M, Livingstone A, Pit S. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
4. Wright J, Caudill R. Remote Treatment Delivery in Response to the COVID-19 Pandemic. Psychother Psychosom 2020;89(3):130-132 [FREE Full text] [doi: 10.1159/000507376] [Medline: 32213775]
5. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
6. Telemedicine: Opportunities and developments in member states: Report on the second global survey on eHealth 2009. World Health Organization. 2010. URL: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2020-12-10]
7. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
8. Downes MJ, Mervin MC, Byrnes JM, Scuffham PA. Telephone consultations for general practice: a systematic review. Syst Rev 2017 Jul 03;6(1):128 [FREE Full text] [doi: 10.1186/s13643-017-0529-0] [Medline: 28673333]
9. Vaona A, Pappas Y, Grewal R, Ajaz M, Majeed A, Car J. Training interventions for improving telephone consultation skills in clinicians. Cochrane Database Syst Rev 2017 Jan 05;1:CD010034 [FREE Full text] [doi: 10.1002/14651858.CD010034.pub2] [Medline: 28052316]
10. Hassounah M, Raheel H, Alhefzi M. Digital Response During the COVID-19 Pandemic in Saudi Arabia. J Med Internet Res 2020 Sep 01;22(9):e19338 [FREE Full text] [doi: 10.2196/19338] [Medline: 32790642]
11. Al-rayes SA, Aldossary H, Aldoukhi E, Alahmedalyousif Z, Aldawood G, Alumran A. The awareness and utilization of 937-telephone health services in Saudi Arabia: Cross-sectional survey study. Inform Med Unlocked 2020;20:100393. [doi: 10.1016/j.imu.2020.100393]
12. Comrey AL, Lee HB. A First Course in Factor Analysis. New York, NY: Psychology Press; 1992.
13. IBM SPSS Statistics for Macintosh. IBM. 2017. URL: https://www.ibm.com/support/pages/node/618179 [accessed 2021-03-11]
14. Leite H, Hodgkinson IR, Gruber T. New development: 'Healing at a distance'—telemedicine and COVID-19. Public Money Manag 2020 Apr 09;40(6):483-485. [doi: 10.1080/09540962.2020.1748855]
15. Knowles E, O'Cathain A, Turner J, Nicholl J. Awareness and use of a new urgent care telephone service, NHS 111: cross-sectional population survey. J Health Serv Res Policy 2014 Oct;19(4):224-230. [doi: 10.1177/1355819614535571] [Medline: 24819379]
16. Hallam L. Access to general practice and general practitioners by telephone: the patient's view. Br J Gen Pract 1993 Aug;43(373):331-335 [FREE Full text] [Medline: 8251217]
17. Hasani SA, Ghafri TA, Al Lawati H, Mohammed J, Al Mukhainai A, Al Ajmi F, et al. The Use of Telephone Consultation in Primary Health Care During COVID-19 Pandemic, Oman: Perceptions from Physicians. J Prim Care Community Health 2020;11:2150132720976480 [FREE Full text] [doi: 10.1177/2150132720976480] [Medline: 33307943]
18. Car J, Sheikh A. Telephone consultations. BMJ 2003 May 03;326(7396):966-969 [FREE Full text] [doi: 10.1136/bmj.326.7396.966] [Medline: 12727771]
19. Lin D, Friedman D, Qiao S, Tam CC, Li X, Li X. Information uncertainty: a correlate for acute stress disorder during the COVID-19 outbreak in China. BMC Public Health 2020 Dec 07;20(1):1867 [FREE Full text] [doi: 10.1186/s12889-020-09952-3] [Medline: 33287780]
20. Escoffery C. Gender Similarities and Differences for e-Health Behaviors Among U.S. Adults. Telemed J E Health 2018 May;24(5):335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]
21. Khatun F, Heywood AE, Hanifi SMA, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
22. Jung S, Kweon H, Kim E, Kim S, Choi J, Cho D. Preference and awareness of telemedicine in primary care patients. Korean J Fam Med 2012 Jan;33(1):25-33 [FREE Full text] [doi: 10.4082/kjfm.2012.33.1.25] [Medline: 22745885]
23. El-Mahalli AA, El-Khafif SH, Al-Qahtani MF. Successes and challenges in the implementation and application of telemedicine in the eastern province of Saudi Arabia. Perspect Health Inf Manag 2012;9:1-27 [FREE Full text] [Medline: 23209455]
24. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]

XSL•FO
RenderX

25. Bujnowska-Fedak M, Grata-Borkowska U. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. Smart Homecare Technol Telehealth 2015 May;3:91-105 [FREE Full text] [doi: 10.2147/shtt.s59498]
26. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-09-15]
27. Norman G, Monteiro S, Salama S. Sample size calculations: should the emperor's clothes be off the peg or made to measure? BMJ 2012 Aug 23;345:e5278. [doi: 10.1136/bmj.e5278] [Medline: 22918496]

Abbreviations

KSA: Kingdom of Saudi Arabia
MOH: Ministry of Health

Edited by C Basch; submitted 03.02.21; peer-reviewed by E van der Velde, E Opoku-Agyemang; comments to author 20.04.21; revised version received 06.05.21; accepted 02.06.21; published 30.07.21.

Please cite as:
Al-Rayes SA, Alumran A, Aljabri D, Aljaffary A, Aldoukhi E, Alahmedalyousif Z, Al Madani R
Public Awareness and Utilization of 937-Telephone Health Services in the Kingdom of Saudi Arabia Before and During the COVID-19 Pandemic: Longitudinal Study
J Med Internet Res 2021;23(7):e27618
URL: https://www.jmir.org/2021/7/e27618
doi:10.2196/27618
PMID:34156963

©Saja A Al-Rayes, Arwa Alumran, Duaa Aljabri, Afnan Aljaffary, Ethar Aldoukhi, Zainab Alahmedalyousif, Reem Al Madani. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online-Delivered Group and Personal Exercise Programs to Support Low Active Older Adults' Mental Health During the COVID-19 Pandemic: Randomized Controlled Trial

Mark R Beauchamp[1], PhD; Ryan M Hulteen[2], PhD; Geralyn R Ruissen[1], MA; Yan Liu[1], PhD; Ryan E Rhodes[3], PhD; Colin M Wierts[1], MSc; Katrina J Waldhauser[1], MA; Samantha H Harden[4], PhD; Eli Puterman[1], PhD

[1]School of Kinesiology, University of British Columbia, Vancouver, BC, Canada

[2]School of Kinesiology, Louisiana State University, Baton Rouge, LA, United States

[3]School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

[4]Department of Human Nutrition, Foods, and Exercise, Virginia Tech, Blacksburg, VA, United States

Corresponding Author:
Mark R Beauchamp, PhD
School of Kinesiology
University of British Columbia
6081 University Blvd
Vancouver, BC, V6T 1Z1
Canada
Phone: 1 604 822 4864
Fax: 1 604 822 6842
Email: mark.beauchamp@ubc.ca

Abstract

Background: In response to the COVID-19 pandemic, experts in mental health science emphasized the importance of developing and evaluating approaches to support and maintain the mental health of older adults.

Objective: The aim of this study was to assess whether a group-based exercise program relative to a personal exercise program (both delivered online) and waitlist control (WLC) can improve the psychological health of previously low active older adults during the early stages of the COVID-19 pandemic.

Methods: The Seniors COVID-19 Pandemic and Exercise (SCOPE) trial was a 3-arm, parallel randomized controlled trial conducted between May and September 2020 in which low active older adults (aged ≥65 years) were recruited via media outlets and social media. After baseline assessments, consented participants were randomized to one of two 12-week exercise programs (delivered online by older adult instructors) or a WLC condition. A total of 241 older adults (n=187 women) provided baseline measures (via online questionnaires), were randomized (n_{group}=80, $n_{personal}$=82, $n_{control}$=79), and completed measures every 2 weeks for the duration of the trial. The trial's primary outcome was psychological flourishing. Secondary outcomes included global measures of mental and physical health, life satisfaction, and depression symptoms.

Results: The results of latent growth modeling revealed no intervention effects for flourishing, life satisfaction, or depression symptoms (*P*>.05 for all). Participants in the group condition displayed improved mental health relative to WLC participants over the first 10 weeks (effect size [ES]=0.288-0.601), and although the week 12 effect (ES=0.375) was in the same direction the difference was not statistically significant (*P*=.089). Participants in the personal condition displayed improved mental health, when compared with WLC participants, in the same medium ES range (ES=0.293-0.565) over the first 8 weeks, and while the effects were of a similar magnitude at weeks 10 (ES=0.455, *P*=.069) and 12 (ES=0.258, *P*=.353), they were not statistically significant. In addition, participants in the group condition displayed improvements in physical health when compared with the WLC (ES=0.079-0.496) across all 12 weeks of the study following baseline. No differences were observed between the personal exercise condition and WLC for physical health (slope *P*=.271).

Conclusions: There were no intervention effects for the trial's primary outcome (ie, psychological flourishing). It is possible that the high levels of psychological flourishing at baseline may have limited the extent to which those indicators could continue to improve further through intervention (ie, potential ceiling effects). However, the intervention effects for mental and physical

XSL•FO
RenderX

health point to the potential capacity of low-cost and scalable at-home programs to support the mental and physical health of previously inactive adults in the COVID-19 pandemic.

Trial Registration: ClinicalTrials.gov NCT04412343; https://clinicaltrials.gov/ct2/show/NCT04412343

(J Med Internet Res 2021;23(7):e30709) doi:10.2196/30709

KEYWORDS

COVID-19; randomized trial; mental health; physical activity

Introduction

As the full scale and impact of the COVID-19 pandemic became evident in early 2020, older adults were identified as being particularly susceptible to severe illness and mortality [1]. National and local governments across the globe subsequently implemented a range of physical distancing mandates, which meant that older adults, in particular, were identified as being at risk of social isolation [2,3]. In direct response, mental health experts emphasized the importance of developing approaches to support and maintain the physical and mental health of older adults during this unprecedented time [2].

One widely scalable, nonpharmacological, and cost-effective approach promoted by the World Health Organization to support mental health during the pandemic corresponds to regular physical activity [4]. Although some correlational studies [5,6], including those focused on older adults [7,8], point to the possibility that regular physical activity may protect against depleted psychological well-being during the pandemic, there has been a distinct absence of experimental studies through which causality might be better ascertained. In this study we sought to examine the efficacy of 2 different types of exercise programs, both delivered online, to support the mental health of previously low active older adults (accumulating ≤60 minutes of moderate-intensity activity per week) within the context of the current COVID-19 pandemic in comparison to participants in a waitlist control (WLC) condition.

Results of previous (pre-COVID-19) experimental research suggest that when older adults exercise in groups with other older adults, led by older adult instructors, and have the opportunity to socially connect after classes, they displayed improvements in group cohesion (ie, they feel more connected) [9], adherence behavior [10], and psychological flourishing [11] when compared with older adults who exercise in classes with middle-aged and younger adults. Other research similarly indicates that when people exercise in group settings, especially within groups that are cohesive, they tend to stick with those programs to a greater extent than when exercising on their own [12]. Given the inability to exercise in community or social settings (eg, fitness/community centers) early in the COVID-19 pandemic, we sought to examine whether a virtually delivered group-based exercise program that sought to promote social connectedness among older adults would derive improved psychological well-being when compared with a personal exercise program (also delivered online) that did not operationalize social connectivity, as well as a WLC condition. The group-based program was informed by the tenets of self-categorization theory [13-16], which indicates that when people share common characteristics (eg, shared identity as older adults) and feel more connected to other group members, they will be more likely to retain their membership of those groups and display greater sense of well-being [17]. The results of a recent meta-analysis of interventions that were designed to foster a sense of social connectivity (and shared social identities) resulted in significant improvements in physical health as well as improvements (in the medium to large effect size [ES] range) in quality of life and cognitive health, as well as reductions in anxiety, depression, and stress [18]. In this trial, we identified psychological flourishing as the a priori primary outcome measure. Psychological flourishing has been identified as an important broad indicator of well-being [19,20], which involves feeling engaged in daily life, optimistic, having a sense of meaning and purpose, and having positive relationships [21]. Flourishing has also been identified as a viable target for intervention [22]. We hypothesized that older adults randomized to the virtual group program would display better well-being (higher levels of psychological flourishing) than those in a personal exercise condition, who in turn would display better well-being than WLCs. As secondary outcomes, we also assessed global measures of mental and physical health, life satisfaction, and depression symptoms. Furthermore, because individuals who live alone may benefit more from a group-based exercise program that fosters social connectivity compared with those who live with others, we investigated whether intervention effects are stronger in those who live alone versus with others. The above hypotheses were pre-registered via the Open Science Framework and ClinicalTrials.gov (see the “Methods” section).

Methods

Trial Design

The Seniors COVID-19 Pandemic and Exercise (SCOPE) study was a prospective, 3-arm, parallel, randomized controlled trial. The corresponding groups underwent a synchronous group-based exercise or an asynchronous personal exercise program (both delivered online), or a WLC condition. The study procedures were approved by the Research Ethics Board of The University of British Columbia, with the design, conduct, and reporting of this study adhering to the Consolidated Standards of Reporting Trials (CONSORT) guidelines [23] (Multimedia Appendix 1). The trial was pre-registered at ClinicalTrials.gov (#NCT04412343) and the Open Science Framework [24].

Participants

Low active older adults (aged ≥65 years) without any medical contraindication that might prevent them from participating in moderate-intensity physical activity were eligible to participate. Additional inclusion criteria included (1) the ability to speak and read English, (2) currently live in Canada, (3) 1 participant

XSL•FO
RenderX

in the study per household, and (4) able to access the internet at home via a personal smartphone, tablet (eg, iPad), or computer (with camera functionality). Activity status was assessed using the Stanford Leisure-Time Categorical Activity Item Version 2.2 (L-CAT; Version 2.2) [25], whereby participants select 1 of 6 descriptive categories ranging from inactive to very active. Consistent with previous use of the L-CAT to screen for low active participants [26], only participants who scored between 1 and 3 were eligible to participate. Item 1 corresponds to activity "no more than once or twice a month," item 2 reflects undertaking "light activities once or twice a week," and item 3 reflects "moderate-intensity activities 3 times per week for 15-20 minutes each time or sport or moderate-intensity activities once per week for 45-60 minutes." As such, item 3 (≤60 minutes of moderate-intensity activity per week) reflects a threshold below current recommendations of 150 minutes of moderate-intensity activity per week for older adults [27]. Prescreening also involved completion of the Physical Activity Readiness Questionnaire for Everyone (PARQ+) and the Electronic Physical Activity Readiness Medical Examination (ePARmed-X+ [28]). If the ePAR-medX+ highlighted that physician approval was required prior to joining the program, the respective individual was informed that this approval was required before they could enroll in the study. Following the initial screening process, informed consent was obtained.

Participants were recruited via social media advertisements (eg, Twitter, Facebook) and news coverage related to the trial (radio, print media), which directed them to the study website. Interested participants were invited to contact the trial coordinator (RH) who scheduled a scripted eligibility screening phone call with a member of the research team. After ascertaining eligibility, interested participants provided informed consent and completed baseline measures for all study measures (ie, demographics plus all health measures) online via Qualtrics (Freedom of Information and Protection of Privacy Act [FIPPA] compliant [29]).

Study Interventions

Participants in the 2 experimental conditions were subsequently directed to a password-protected and secured web platform housed by the first author's institution (ie, Canvas). This platform provided access to the appropriate exercise programs and intervention materials. Individuals randomized to undergo the group-based exercise program received an adapted version of the group program that was previously implemented for older adults for in-person exercise classes [10]. Specifically, participants had the opportunity to take part in group exercise classes delivered via an online communications platform (ie, Zoom housed within Canvas) by older adult exercise instructors (n_{men}=3, n_{women}=4; mean age 68.29 [SD 8.90] years) that were employed at a local community center and had considerable experience delivering older adult exercise classes. Classes were offered 7 days a week at 9 am Pacific Standard Time (12 noon Eastern Standard Time, 1 pm Atlantic Standard Time), and lasted approximately 50-60 minutes. Classes included a warm-up component, moderate-intensity exercises as the core component of the class, and a cool-down period, and were designed specifically for older adults to include strength, flexibility, balance, and aerobic components. Consistent with international guidelines for weekly physical activity by older adults (150 minutes of moderate-to-vigorous intensity physical activity [27]), participants were encouraged to attend at least three classes each week. Classes were hosted on Zoom by a trained research assistant (RH, GR, and CW) who provided technical assistance to ensure that the classes were accessible to participants.

Instructors were provided with autonomy support [30], whereby they could choose the exercises included in each class; however, to ensure sufficient support, instructors were directed to ensure that all exercises could be completed in the home environment with minimal need for equipment (resistance bands were sent, by postal mail, to all participants to facilitate strength-based exercises). Instructors were encouraged to use language in their classes to foster a sense of "us" and that "we're in this together" (ie, to develop a sense of social identity and connectedness even though classes were delivered virtually). At the end of classes, participants had the opportunity to connect in small groups (via Zoom breakout groups) to socially connect over a beverage (coffee, water) from their own homes. If participants missed the live class, they could access a recording of that class in their own time. Participants were also sent, by postal mail, a program t-shirt to foster a sense of distinctiveness [31].

The same older adult instructors that delivered the group exercise classes also delivered the personal exercise classes (they were blind to the trial hypotheses). Classes in the personal exercise condition were matched for frequency, duration, intensity, and content with the group exercise classes but, in this instance, instructors used language during the classes that referred to themselves as each participant's personal trainer/coach, and language directed to the individual and not any group. That is, no sense of "groupness" or "shared social identity/connectivity" was primed. Classes were pre-recorded and accessed via Canvas, which meant that they could be accessed any time in the day of the participants' choosing. Participants in this condition did not have the opportunity to interact with other program participants after classes had ended, and did not receive the same program t-shirts designed to foster a sense of group distinctiveness.

Older adults randomized to the control condition were asked to go about their daily lives for the duration of the 12-week trial. They were asked to complete the same questionnaires (and were remunerated in the same way) as those in the other 2 conditions. At the end of the 12-week trial, participants in this condition were provided with access to (and supports associated with) the personal exercise program described above.

Participants in all 3 conditions were sent questionnaires related to the trial's measures (see below) for completion (via Qualtrics) at the end of weeks 2, 4, 6, 8, 10, and 12. In return for survey completion, at each time point, participants were provided Can $10 (US $8) (Can $70 [US $56] total; baseline plus 6 follow-up assessments). Participants also received up to Can $50 (US $40) if any costs were incurred for obtaining medical clearance from their respective family doctor.

XSL•FO
RenderX

Measures

The primary outcome measure was psychological flourishing, which we assessed using Diener and colleagues' [21] 8-item measure. Exemplar items include "I lead a purposeful and meaningful life" and "I am engaged and interested in my daily activities," with all items anchored on a 7-point Likert scale from "strongly disagree" (1) to "strongly agree" (7). Responses to the flourishing scale demonstrated acceptable reliability, with Cronbach α values ranging from .90 to .94 across the 7 time points. Secondary outcomes included global measures of mental and physical health, life satisfaction, and depression symptoms. Global mental and physical health were assessed using separate 1-item measures developed by Hays and colleagues [32]. Specifically, mental health was assessed using the item "In general, would you say your MENTAL OR EMOTIONAL HEALTH is excellent, very good, good, fair, or poor?," while physical health was assessed using the item "In general, would you say your PHYSICAL HEALTH is excellent, very good, good, fair, or poor?," with each item anchored from "poor" (1) to "excellent" (5). Life satisfaction was assessed using the 1-item question by Fleeson [33] that asked participants "Using a scale from 0 to 10 where 0 means 'the worst possible life overall' and 10 means 'the best possible life overall', how would you rate your life overall these days?." Depression symptoms were assessed using the 10-item Center for Epidemiologic Studies Depression Scale (CES-D) [34] that asked participants to report the frequency of depression symptoms over the past week. Exemplar items include "I felt that everything I did was an effort" and "I felt depressed," with all items anchored by "Rarely or none of the time (less than 1 day)" and "Most or all of the time (5-7 days)." Responses to the CES-D demonstrated acceptable reliability, with Cronbach α values ranging from .83 to .87 across the 7 time points. In addition, participants completed measures of chronic health conditions [35], as well as demographic measures that recorded sex and gender, age, type of dwelling, ethnicity, sexual orientation, smoking behavior, height, weight, education level, household income level, employment status, marital status, and living situation (ie, living alone versus with others). Participant engagement in each of the physical activity programs, as a measure of program adherence, was operationalized via the data analytics for each individual within the Canvas platform, where each of the classes/sessions were provided. As a manipulation check, participants were considered to have attended a class/session if they recorded 10 or more minutes of class/session access.

Sample Size Calculation

To account for interdependence among observations (ie, multiple observations within the same participant), we conducted the power analysis using Optimal Design Software [36]. On the basis of 7 observations (baseline, plus weeks 2, 4, 6, 8, 10, and 12), a total sample size of 527 was identified as necessary to detect a small ES (in psychological flourishing) of δ=0.25, with intraclass correlation coefficient set at 0.05, power (1–β) at 80%, and α at .05 with 7 time points. To account for a study attrition rate of 10% (over the course of the study), a sample size of 600 was considered sufficient to examine the latent growth models (LGMs) proposed in this trial.

Randomization and Blinding

Participants were stratified to ensure equal distribution of men and women across conditions. Sequence generation was completed separately for men and women using the Research Randomizer [37] tool for researchers, with blocks of 3 unique numbers (1, 2, and 3) that designated 1 of the 3 randomization groups. A researcher external to the project team generated the sequence and remained blind to participant allocation. Participants were randomized in the order they completed baseline surveys. Although the trial coordinator (RH) was aware of condition assignment (following randomization), there were no experimenter or investigator expectancy effects related to the mental health outcome measures as all assessments occurred online (ie, online questionnaires). Once baseline measures were completed, the trial coordinator contacted each participant to inform him/her of the condition assignment as a result of the trial's randomization procedures.

Changes to the Trial

On July 17, 2020, study recruitment was terminated for several reasons. First, during our recruitment window (June and July 2020), there were some major global events (eg, protests, riots, and political events) which limited our ability to get the word out via the media. Although we had some success with media recruitment (national radio, TV, print media), our recruitment did not have the anticipated reach. Second, we wanted to keep the recruitment window similar across participants: we anticipated that the experiences of older adults early in the pandemic would be notably different to those experienced months later (eg, summer versus winter, along with geopolitical changes across time). Closing new enrollments at that point meant that all participants commenced the study at the same time of the year (within a 5-week window) and had started at approximately the same time as one another in relation to the unfolding pandemic. Third, although a sample of 600 was required to detect a small effect, under the same parameters as originally presented (ie, intraclass correlation coefficient set at 0.05, power [1–β] at 80%, and α at .05 with 7 time points including baseline, weeks 2, 4, 6, 8, 10, and 12), a sample of 209 was required to detect a small-to-medium effect of δ=0.40 and a sample of 134 was required to detect a medium-sized effect of δ=0.50. Thus, even after accounting for the original attrition rate of 10%, a sample of 241 was deemed sufficient to detect small-to-medium and medium ESs of 0.40 and 0.50, respectively. As there were no feasibility/efficacy data to sufficiently gauge the size of an intervention effect in the context of a pandemic, we felt it was appropriate to cease new enrollments, while acknowledging that the trial would not be sufficiently powered to detect small effects, but would be well powered to detect small-to-medium and medium-sized effects. More information on these changes can be accessed here [38].

We originally also sought to examine whether any intervention effects might be more pronounced among those with lower mental health at baseline [24]. Unfortunately, we were precluded from conducting these subgroup analyses due to the resultant small sample size and instability of parameter estimates. For example, research using the CES-D has identified a threshold score of 10 or more as indicative of depression symptomology

XSL•FO
RenderX

[34]. In our study, 76 older adults met this criterion on the basis of their baseline scores.

Statistical Analyses

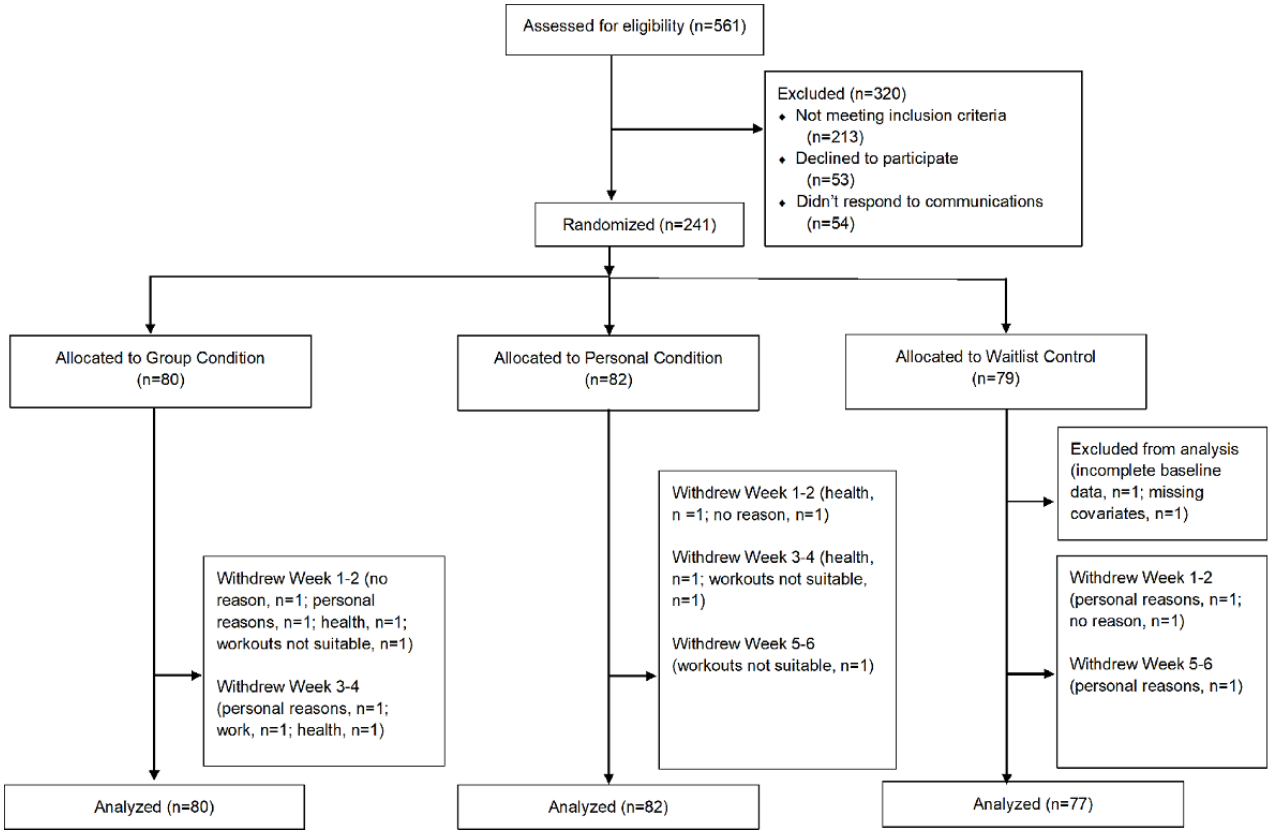
We conducted our main data analyses for the 5 outcome variables using latent growth modeling based on a structural equation modeling framework, including all randomized participants (intention-to-treat analyses), using the Mplus version 7.4 software [39] with maximum likelihood robust estimation (Multimedia Appendix 2). As the data were collected on multiple occasions over 12 weeks following baseline assessments, we tested both linear and nonlinear trends. First, we conducted an unconditional growth model, and compared linear (Multimedia Appendix 3) and quadratic (Multimedia Appendix 4) growth models, and determined the optimal model through commonly used model fit indices. This corresponded to the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). The criteria for evaluating model fit was designated with CFI values over 0.90, and RMSEA and SRMR values less than 0.08 [40,41]. Quadratic models were utilized to take account of nonlinear growth trends. Second, to test the hypothesized treatment effects, we included the intervention conditions (personal exercise versus group exercise) in the analysis, and controlled for the effects of covariates, including sex, age, living situation (ie, "alone" versus "with others"), and chronic health conditions. In light of our a priori hypothesis that living status would moderate the intervention effects, we included the interaction of living situation and experimental conditions in this step. We computed ESs at each time point using Feingold's approach [42-44] for growth modeling (equivalent to Cohen *d*).

Results

Overview

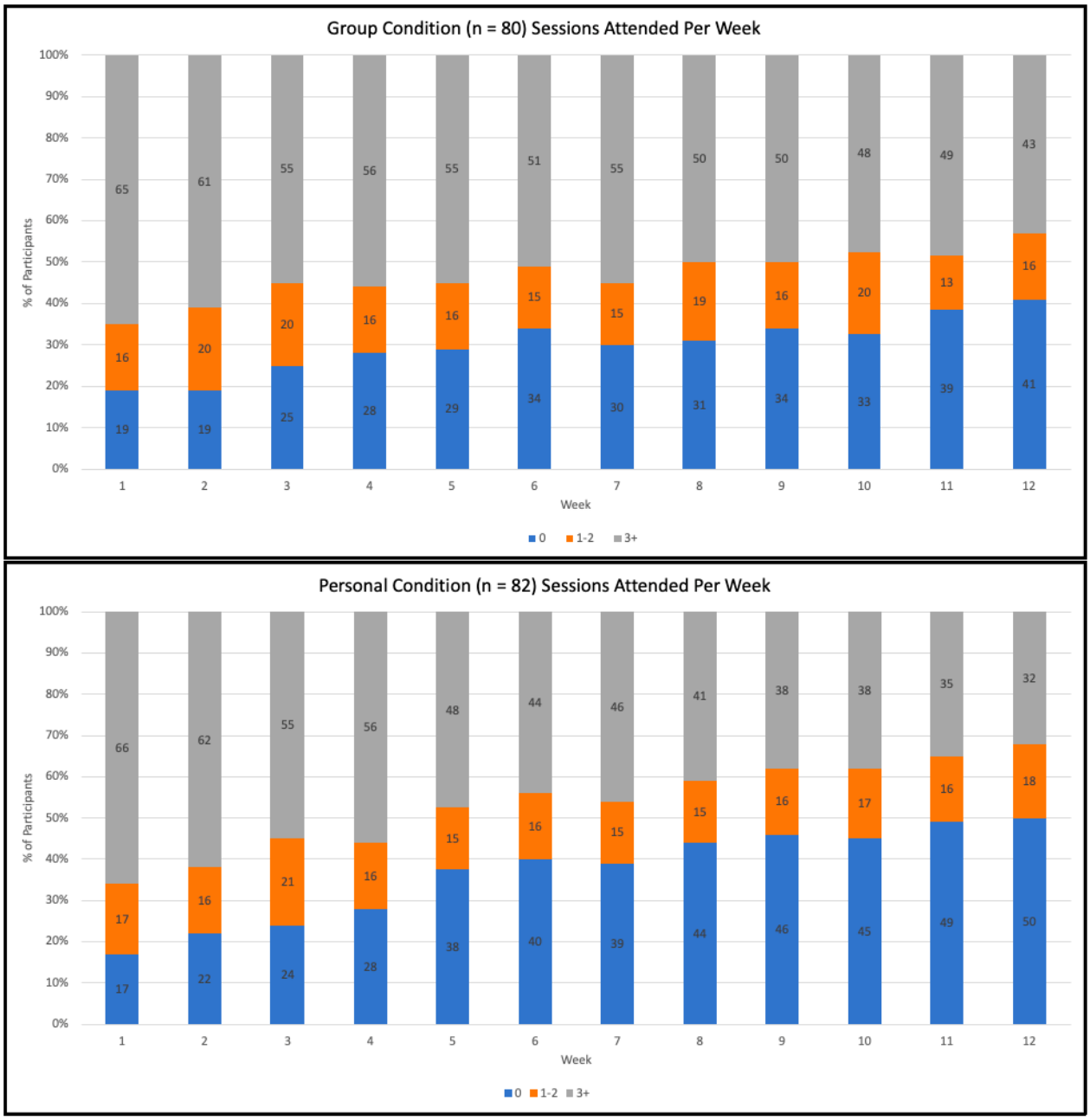
Five hundred and sixty-one individuals were screened and, based on eligibility, 241 adults aged 65-94 (mean age 73.03 [SD 5.42] years) enrolled between May 23 and July 12, 2020 (Figure 1). Descriptive statistics for the sociodemographic factors are presented in Multimedia Appendix 5. There were no differences between groups at baseline (as indicated by the nonsignificant intercepts in Multimedia Appendices 6-10) with regard to any of the 5 dependent measures assessed in the study. Correlations among the study variables at each time point are presented in Multimedia Appendices 11-17. Exercise session attendance for the 2 experimental conditions across the 12 weeks of the trial is presented in Figure 2.

Figure 1. CONSORT flow diagram.



XSL•FO
RenderX

Figure 2. Program attendance in the two experimental conditions across the 12-week trial.



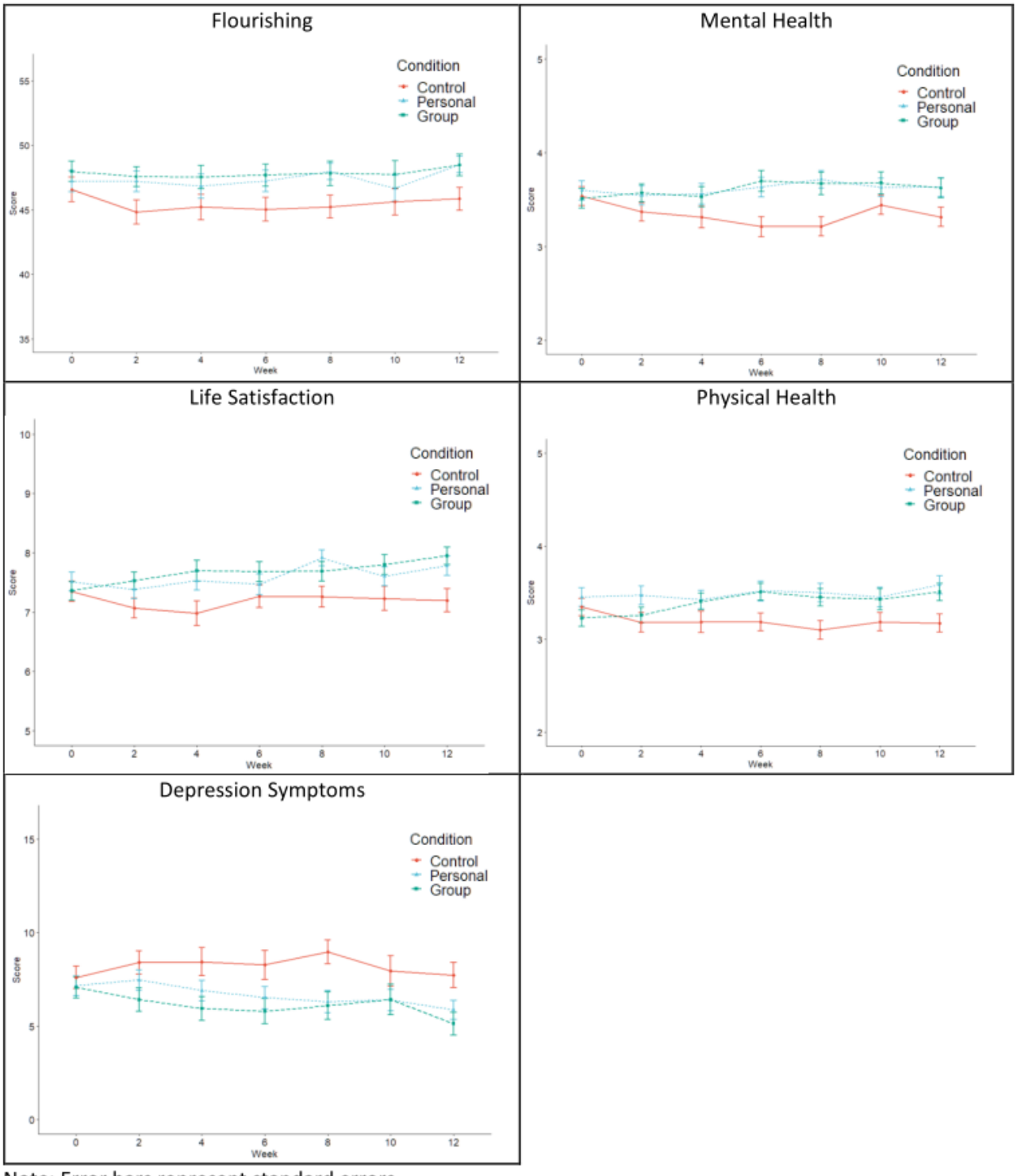
Intervention Groups Versus Control Condition on Study Outcomes

Flourishing

An LGM that accounted for quadratic change displayed good model fit (CFI=0.967, RMSEA=0.048, SRMR=0.067) for flourishing (Multimedia Appendix 6). There were no differences between the 2 intervention conditions and control condition at baseline (as denoted by nonsignificant intercepts; Multimedia Appendix 6 and Figure 3). Older adults living with others versus alone (intercept estimate=3.562, SE=1.171, *P*=.030, 95% CI 0.197-6.926) and those with fewer chronic health conditions (intercept estimate=–0.977, SE=0.205, *P*<.001, 95% CI –1.379 to –0.576) displayed higher levels of flourishing at baseline. After controlling for covariates, there was no significant intervention effect for either personal or group exercise conditions in comparison to the WLC and a nonsignificant intervention condition by living situation interaction (as denoted by nonsignificant slopes; Multimedia Appendix 6).

XSL•FO
RenderX

Figure 3. Trajectories for mental health outcomes over the course of the trial.



Note: Error bars represent standard errors.

Mental Health

An LGM that accounted for quadratic change displayed good model fit (CFI=0.992, RMSEA=0.025, SRMR=0.031) for mental health (Multimedia Appendix 7). There were no differences between the 2 intervention conditions and control condition at baseline (as denoted by nonsignificant intercepts; Multimedia Appendix 7). With regard to covariates, men reported better mental health at baseline (intercept estimate=0.286, SE=0.132, *P*=.030, 95% CI 0.027-0.544), as did those who were older (intercept estimate=0.044, SE=0.010, *P*<.001, 95% CI 0.024-0.064), and those with fewer chronic health conditions (intercept estimate=–0.116, SE=0.027, *P*<.001, 95% CI –0.169 to –0.064). After controlling for covariates, both the personal exercise (slope estimate=0.291, SE=0.123, *P*=.010, 95% CI 0.050-5.32) and group exercise (slope estimate=0.282, SE=0.099, *P*=.004, 95% CI 0.088-0.476) conditions displayed improved mental health when compared with the WLC condition (Multimedia Appendix 7 and Figure 3). The significant quadratic effects for both intervention conditions in relation to the WLC illustrate differences in curvature of the slopes when compared

XSL•FO
RenderX

with the control condition (Multimedia Appendix 7 and Figure 3). The differences in mental health between the personal exercise condition and the WLC were in the medium ES range (ES=0.293-0.565) over the first 8 weeks and although the effects were of a similar magnitude at weeks 10 (ES=0.455, *P*=.069) and 12 (ES=0.258, *P*=.353), they were not statistically significant. The differences in mental health between the group condition and WLC were in the same medium ES range over the first 10 weeks (ES=0.288-0.601), and although the week 12 effect (ES=0.375) was in the same direction, the difference was not statistically significant (*P*=.089). The intervention condition by living situation interactions were nonsignificant (Multimedia Appendix 7).

Physical Health

An LGM that accounted for linear change displayed good model fit (CFI=0.977, RMSEA=0.037, SRMR=0.037) for self-reported physical health (Multimedia Appendix 8). Model fit did not improve by modeling quadratic change. With regard to covariates, participants who were older (intercept estimate=0.033, SE=0.009, *P*<.001, 95% CI 0.015-0.050), and those with fewer chronic health conditions (intercept estimate=–0.116, SE=0.025, *P*<.001, 95% CI –0.168 to –0.071) reported better physical health at baseline. After controlling for covariates, participants in the group condition displayed improvements in physical health when compared with the WLC (slope estimate=0.063, SE=0.030, *P*<.001, 95% CI 0.004-0.121), with the ESs observed in the small to medium range (ES=0.079-0.496) across all 12 weeks of the study following baseline (Multimedia Appendix 8 and Figure 3). After controlling for covariates, the difference between the personal exercise condition and WLC was not significant. None of the condition by living situation interactions were significant (Multimedia Appendix 8).

Life Satisfaction

An LGM that accounted for linear change displayed good model fit (CFI=0.947, RMSEA=0.059, SRMR=0.067) for life satisfaction (Multimedia Appendix 9). Model fit did not improve by modeling quadratic change. With regard to covariates, men reported greater life satisfaction at baseline (intercept estimate=0.623, SE=0.191, *P*=.001, 95% CI 0.250-0.997), as did those who were older (intercept estimate=0.039, SE=0.014, *P*=.007, 95% CI 0.011-0.068) and those with fewer chronic health conditions (intercept estimate=–0.177, SE=0.039, *P*<.001, 95% CI –0.252 to –0.101). After controlling for covariates, there was no significant intervention effect for either personal or group exercise condition in comparison to the WLC and the condition by living situation interactions were also nonsignificant (as denoted by nonsignificant slopes; Multimedia Appendix 9 and Figure 3).

Depression Symptoms

An LGM that accounted for quadratic change displayed good model fit (CFI=0.992, RMSEA=0.024, SRMR=0.024) for depression symptoms (Multimedia Appendix 10). With regard to covariates, women reported less depression symptoms at baseline (intercept estimate=–1.454, SE=0.639, *P*=.023, 95% CI –2.707 to –0.202), as did those who were older (intercept estimate=–0.133, SE=0.050, *P*=.008, 95% CI –0.232 to –0.035]), while those with more chronic health conditions reported higher depression symptoms (intercept estimate=0.703, SE=0.150, *P*<.001, 95% CI 0.408-0.997) at baseline (Multimedia Appendix 10). There were no significant intervention effects and none of the condition by living status interactions were significant (Multimedia Appendix 10 and Figure 3).

Discussion

Principal Findings

The overall purpose of this study was to test the efficacy of 2 physical activity programs to support previously low active older adults' psychological and physical well-being early in the COVID-19 pandemic. Both physical activity interventions were delivered online, with one designed to foster a sense of social connectivity, and the other designed to support independent physical activity, and compared against a control condition. After displaying comparable levels of program adherence over the first 4 weeks of the trial, participants in the group program displayed improved adherence compared with those in the personal exercise program; over the last 4 weeks the proportion of participants attending 3 or more sessions per week was 10% or more in the group condition than in the personal condition (Figure 1). Despite this, there were no intervention effects for either condition, in relation to the trial's primary outcome, psychological flourishing, or measures of life satisfaction and depression symptoms. Both intervention conditions did, however, display significant intervention effects (in the medium ES range) for a global/omnibus measure of mental health when compared with the control condition. In addition, participants in the group exercise condition demonstrated significant intervention effects, again in the medium ES range, for self-reported physical health when compared with controls.

Early in the pandemic, older adults were identified as being particularly at risk of isolation and depleted well-being [2,3], and as such represented the focus of intervention in this trial. As the first few months of the pandemic progressed, the results of large-scale epidemiology studies in North America revealed that, perhaps contrary to initial expectations, older adults displayed the lowest prevalence of psychological distress within any age group (adults aged 18-29 displayed the highest levels of distress) [45]. With this in mind, it is notable that older adults who were screened for eligibility and enrolled in the study displayed generally good psychological health at baseline (ie, high mean levels of flourishing and low mean levels of depression symptoms). Indeed, it is likely that the high levels of psychological flourishing reported for the overall sample at baseline may have limited the extent to which those indicators could continue to improve further through intervention (ie, potential ceiling effects). It is also noteworthy that psychological flourishing represents a multicomponent indicator of well-being [22] that includes aspects such as having positive relationships, feelings of competence, and having meaning and purpose in life [21], and so it is certainly conceivable that the interventions tested in this study were not sufficiently potent to improve such a broad multicomponent indicator. The same could also be said of life satisfaction, which also displayed null effects in this trial.

XSL•FO
RenderX

Nevertheless, the trial did result in significant intervention effects for both experimental conditions in relation to global indicators of mental health when compared with the control condition. In addition, it is noteworthy that involvement in the group condition resulted in medium-sized effects in self-reported physical health. While the trial resulted in intervention effects for these 2 measures, and null effects for the other 3 mental health measures (flourishing, life satisfaction, and depression symptoms), it is notable that the trajectories, or patterns, of all 5 of the study measures (Figure 3) are directly comparable to one another (with depression symptoms displaying mirrored trajectories; with lower depression symptoms reflecting better psychological health). We also examined whether living situation might moderate the effects of the interventions in relation to the targeted study outcomes, and hypothesized a priori that the physical activity programs (in particular, the group-based program) would be more effective in contributing to participant well-being for those who live alone than with others. No significant condition by living situation interactions were observed, which suggests that no differential intervention effects occurred based on participants' living status.

From a knowledge translation perspective several findings are worthy of note. First, the 2 intervention conditions were directly matched in relation to the content and frequency of classes/sessions, with the same older adult instructors delivering the classes across conditions. Although the personal exercise condition had built-in flexibility, whereby participants could access the classes/sessions during times of their own choosing, the adherence data indicate that the opportunity to exercise alongside other older adults (in an online group-based program) provided an added draw to sustain their involvement. Indeed, while the adherence levels were directly comparable across conditions for the first 4 weeks of the trial, over the remaining 8 weeks (likely as participants become more familiar with one another) those in the group program displayed improved adherence behaviors. Although older adults may not accrue the same quality of connections with other class members that occurs within more typical in-person groups, the adherence data indicated that online groups can act to substantively retain older adults' involvement in physical activity programs (at least in the context of a global pandemic). Second, the programs delivered in this trial were designed in such a way that all exercises could be completed in the home environment with minimal need for equipment. Thus, provided that participants had access to the internet at home via a personal smartphone, tablet (eg, iPad), or computer, there were no barriers to participation. We recognize that some older adults face digital inequalities that limit their access to the internet (and programs delivered via the internet) [46]. Nevertheless, with the proportion of older adults who have access to the internet doubling between 2007 and 2016 in Canada (from 32% to 68%) [47], and with trends expected to improve further [47], programs such as those delivered in this trial have considerable potential to be delivered to older adults, either in circumstances such as the current COVID-19 pandemic, or in other contexts such as living in remote or rural communities. Such online programs also have considerable potential to be delivered at scale.

Despite the contributions of the study, limitations must be acknowledged. The most substantive limitation corresponds to the deviation from the initial target sample size (as described in the "Methods" section). Although we utilized social media (Facebook, Twitter) to recruit participants, we found that the overwhelming majority of participants, most likely due to the nature of the older adult demographic, were alerted to the study via news reports in the national/provincial press. We had good uptake from a handful of news reports/stories (to facilitate recruitment), but were unable to secure our targeted sample size. Nevertheless, our eventual sample size (n=241) may well have precluded us from detecting significant between-group effects within the LGMs (smaller sample sizes tend to produce increased standard errors, thus reducing statistical power sensitivity to detect between-group differences). The second limitation corresponded to our measure of physical health. Because of the physical distancing recommendations that existed at the time of conducting the study, it was not possible to conduct in-person assessments of physical health. As such, we utilized a self-report global measure of physical health, which could at best be described as a proxy for actual physical health. As a final limitation, we originally planned to examine whether the effects would be more pronounced for those who displayed worsened mental health at baseline; however, we were precluded from conducting this subgroup analyses due to the small number of participants who displayed identifiably depleted levels of mental health at baseline (eg, CES-D scores ≥10 [34]). Nevertheless, despite these limitations, the study represents one of the few randomized trials to evaluate the efficacy of physical activity interventions during the COVID-19 pandemic, with the findings revealing that virtually delivered interventions are feasible and, when delivered in a group setting, can aid the retention of previously low active older adults. In addition, the results provided some indication that both physical activity programs showed improvements in mental health when compared with control participants, which represents a notable outcome for older adults in the current COVID-19 pandemic.

Conclusions

In response to calls from mental health experts [2] early in the COVID-19 pandemic to design and implement programs to support the mental and physical health of older adults, we sought to examine the efficacy of 2 interventions through the implementation of a pre-registered randomized controlled trial. Both interventions were delivered online, to support previously low active older adults engaging in physical activity while ensuring that they maintained physical distancing (as part of prevailing government pandemic-related public health mandates). Although no significant intervention effects resulted in relation to the trial's primary outcome (ie, psychological flourishing), the intervention effects for both the group and personal conditions in relation to mental and physical health (in the medium ES range) point to the capacity of low-cost and scalable at-home programs, delivered online, to support older adults' well-being in the COVID-19 pandemic, as well as other remote or hard-to-reach rural settings.

XSL•FO
RenderX

Acknowledgments

We acknowledge George Gao, Nathan Sanghe, Joseph O'Rourke, Taha Elramly, and Sophie Wensel for their contributions to recruitment and screening of participants. We also thank Kaitlin Crawford for her technical assistance associated with hosting the exercise classes on Zoom, and Jeff Longland and Craig Thompson for their assistance with procuring the data analytics from Canvas to produce the program attendance data. Funding for this research was provided by the Canadian Institutes of Health Research (Grant # PJT 169211).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1256 KB - jmir_v23i7e30709_app1.pdf]

Multimedia Appendix 2
Mplus code.
[DOCX File , 18 KB - jmir_v23i7e30709_app2.docx]

Multimedia Appendix 3
Latent growth modeling path diagram (modeling linear change) for explanatory variables in relation to the putative dependent measure over the 12-week trial.
[DOCX File , 51 KB - jmir_v23i7e30709_app3.docx]

Multimedia Appendix 4
Latent growth modeling path diagram (modeling quadratic change) for explanatory variables in relation to the putative dependent measure over the 12-week trial.
[DOCX File , 59 KB - jmir_v23i7e30709_app4.docx]

Multimedia Appendix 5
Participant demographic information.
[DOCX File , 18 KB - jmir_v23i7e30709_app5.docx]

Multimedia Appendix 6
Latent growth model (accounting for quadratic change) for psychological flourishing.
[DOCX File , 16 KB - jmir_v23i7e30709_app6.docx]

Multimedia Appendix 7
Latent growth model (accounting for quadratic change) for mental health.
[DOCX File , 17 KB - jmir_v23i7e30709_app7.docx]

Multimedia Appendix 8
Latent growth model (accounting for linear change) for physical health.
[DOCX File , 16 KB - jmir_v23i7e30709_app8.docx]

Multimedia Appendix 9
Latent growth model (accounting for linear change) for life satisfaction.
[DOCX File , 15 KB - jmir_v23i7e30709_app9.docx]

Multimedia Appendix 10
Quadratic latent growth model for depressive symptomology.
[DOCX File , 19 KB - jmir_v23i7e30709_app10.docx]

Multimedia Appendix 11
Baseline correlations among study variables.
[DOCX File , 18 KB - jmir_v23i7e30709_app11.docx]

XSL•FO
RenderX

Multimedia Appendix 12
Correlations among study variables at week 2.
[DOCX File , 18 KB - jmir_v23i7e30709_app12.docx]

Multimedia Appendix 13
Correlations among study variables at week 4.
[DOCX File , 18 KB - jmir_v23i7e30709_app13.docx]

Multimedia Appendix 14
Correlations among study variables at week 6.
[DOCX File , 18 KB - jmir_v23i7e30709_app14.docx]

Multimedia Appendix 15
Correlations among study variables at week 8.
[DOCX File , 18 KB - jmir_v23i7e30709_app15.docx]

Multimedia Appendix 16
Correlations among study variables at week 10.
[DOCX File , 18 KB - jmir_v23i7e30709_app16.docx]

Multimedia Appendix 17
Correlations among study variables at week 12.
[DOCX File , 22 KB - jmir_v23i7e30709_app17.docx]

References

1. Jowell A, Carstensen LL, Barry M. A life-course model for healthier ageing: lessons learned during the COVID-19 pandemic. Lancet Healthy Longev 2020 Oct;1(1):e9-e10. [doi: 10.1016/s2666-7568(20)30008-8]
2. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
3. Banerjee D. The impact of Covid-19 pandemic on elderly mental health. Int J Geriatr Psychiatry 2020 Dec;35(12):1466-1467 [FREE Full text] [doi: 10.1002/gps.5320] [Medline: 32364283]
4. World Health Organization. #HealthyAtHome - Physical Activity.: World Health Organization URL: https://www.who.int/news-room/campaigns/connecting-the-world-to-combat-coronavirus/healthyathome/healthyathome---physical-activity [accessed 2021-06-21]
5. Jacob L, Tully MA, Barnett Y, Lopez-Sanchez GF, Butler L, Schuch F, et al. The relationship between physical activity and mental health in a sample of the UK public: A cross-sectional study during the implementation of COVID-19 social distancing measures. Ment Health Phys Act 2020 Oct;19:100345 [FREE Full text] [doi: 10.1016/j.mhpa.2020.100345] [Medline: 32834833]
6. Faulkner G, Rhodes RE, Vanderloo LM, Chulak-Bozer T, O'Reilly N, Ferguson L, et al. Physical activity as a coping strategy for mental health due to the COVID-19 virus: a potential disconnect among Canadian adults? Front Commun 2020 Sep 24;5:571833. [doi: 10.3389/fcomm.2020.571833]
7. Callow DD, Arnold-Nedimala NA, Jordan LS, Pena GS, Won J, Woodard JL, et al. The mental health benefits of physical activity in older adults survive the COVID-19 pandemic. Am J Geriatr Psychiatry 2020 Oct;28(10):1046-1057 [FREE Full text] [doi: 10.1016/j.jagp.2020.06.024] [Medline: 32713754]
8. Creese B, Khan Z, Henley W, O'Dwyer S, Corbett A, Vasconcelos Da Silva M, et al. Loneliness, physical activity, and mental health during COVID-19: a longitudinal analysis of depression and anxiety in adults over the age of 50 between 2015 and 2020. Int Psychogeriatr 2021 May;33(5):505-514 [FREE Full text] [doi: 10.1017/S1041610220004135] [Medline: 33327988]
9. Beauchamp MR, Liu Y, Dunlop WL, Ruissen GR, Schmader T, Harden SM, et al. Psychological mediators of exercise adherence among older adults in a group-based randomized trial. Health Psychol 2021 Mar;40(3):166-177. [doi: 10.1037/hea0001060] [Medline: 33630638]
10. Beauchamp MR, Ruissen GR, Dunlop WL, Estabrooks PA, Harden SM, Wolf SA, et al. Group-based physical activity for older adults (GOAL) randomized controlled trial: Exercise adherence outcomes. Health Psychol 2018 May;37(5):451-461. [doi: 10.1037/hea0000615] [Medline: 29698019]

XSL•FO
RenderX

11. Ruissen GR, Liu Y, Schmader T, Lubans DR, Harden SM, Wolf SA, et al. Effects of group-based exercise on flourishing and stigma consciousness among older adults: Findings from a randomised controlled trial. Appl Psychol Health Well Being 2020 Jul;12(2):559-583. [doi: 10.1111/aphw.12197] [Medline: 32149456]
12. Burke S, Carron A, Eys M, Ntoumanis N, Estabrooks P. Group versus individual approach? A meta-analysis of the effectiveness of interventions to promote physical activity. Sport Exerc Psychol Rev 2006;2:19-35.
13. Haslam C, Jetten J, Cruwys T, Dingle GA, Haslam SA. The New Psychology of Health: Unlocking the Social Cure. London, UK: Routledge; 2018.
14. Turner JC. Social categorization and the self-concept: A social cognitive theory of group behavior. In: Lawler EJ, editor. Advances in Group Processes: Theory and Research. Greenwich, UK: JAI Press; 1985:72-81.
15. Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS. Rediscovering the Social Group: A Self-Categorization Theory. Oxford and New York: Basil Blackwell; 1987.
16. Turner JC, Oakes PJ, Haslam SA, McGarty C. Self and collective: Cognition and social context. Pers Soc Psychol Bull 2016 Jul 02;20(5):454-463. [doi: 10.1177/0146167294205002]
17. Beauchamp MR, Rhodes RE. A group-mediated approach to precision medicine-social identification, prevention, and treatment. JAMA Psychiatry 2020 Jun 01;77(6):555-556. [doi: 10.1001/jamapsychiatry.2020.0024] [Medline: 32101268]
18. Steffens NK, LaRue CJ, Haslam C, Walter ZC, Cruwys T, Munt KA, et al. Social identification-building interventions to improve health: A systematic review and meta-analysis. Health Psychol Rev 2021 Mar;15(1):85-112. [doi: 10.1080/17437199.2019.1669481] [Medline: 31530154]
19. VanderWeele TJ, McNeely E, Koh HK. Reimagining health-flourishing. JAMA 2019 May 07;321(17):1667-1668. [doi: 10.1001/jama.2019.3035] [Medline: 30933213]
20. VanderWeele TJ, Trudel-Fitzgerald C, Allin P, Farrelly C, Fletcher G, Frederick DE, et al. Current recommendations on the selection of measures for well-being. Prev Med 2020 Mar 01;133:106004 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106004] [Medline: 32006530]
21. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: Short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2009 May 28;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]
22. VanderWeele TJ. On the promotion of human flourishing. Proc Natl Acad Sci U S A 2017 Aug 01;114(31):8148-8156 [FREE Full text] [doi: 10.1073/pnas.1702996114] [Medline: 28705870]
23. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
24. Beauchamp M, Hulteen R, Ruissen G, Liu Y, Rhodes R, Puterman E. The Seniors COvid-19 Pandemic and Exercise (SCOPE) Study: A Randomized Controlled Trial. 2020 May 25. URL: https://osf.io/xw9gm [accessed 2020-05-24]
25. Kiernan M, Schoffman DE, Lee K, Brown SD, Fair JM, Perri MG, et al. The Stanford Leisure-Time Activity Categorical Item (L-Cat): A single categorical item sensitive to physical activity changes in overweight/obese women. Int J Obes (Lond) 2013 Dec;37(12):1597-1602 [FREE Full text] [doi: 10.1038/ijo.2013.36] [Medline: 23588625]
26. Toledo MJ, Hekler E, Hollingshead K, Epstein D, Buman M. Validation of a smartphone app for the assessment of sedentary and active behaviors. JMIR Mhealth Uhealth 2017 Aug 09;5(8):e119 [FREE Full text] [doi: 10.2196/mhealth.6974] [Medline: 28793982]
27. Ross R, Chaput J, Giangregorio LM, Janssen I, Saunders TJ, Kho ME, et al. Canadian 24-hour movement guidelines for adults aged 18-64 years and adults aged 65 years or older: An integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab 2020 Oct;45(10 (Suppl. 2)):S57-S102 [FREE Full text] [doi: 10.1139/apnm-2020-0467] [Medline: 33054332]
28. Warburton D, Jamnik V, Bredin S, Gledhill N. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+): Summary of consensus panel recommendations. Health Fit J Canada 2011;4:3-23. [doi: 10.14288/hfjc.v4i2.103]
29. Government of British Columbia. Freedom of Information and Protection of Privacy Act. 1996. URL: https://www.bclaws.gov.bc.ca/civix/document/id/complete/statreg/96165_00 [accessed 2020-06-20]
30. Mageau GA, Vallerand RJ. The coach-athlete relationship: A motivational model. J Sports Sci 2003 Nov;21(11):883-904. [doi: 10.1080/0264041031000140374] [Medline: 14626368]
31. Beauchamp MR. Promoting exercise adherence through groups: A self-categorization theory perspective. Exerc Sport Sci Rev 2019 Jan;47(1):54-61. [doi: 10.1249/JES.0000000000000177] [Medline: 30334854]
32. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-880 [FREE Full text] [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
33. Fleeson W. The quality of American life at the end of the century. In: Brim OG, Ryff CD, Kessler RC, editors. How Healthy Are We?: A National Study of Well-Being at Midlife. Chicago, IL: University of Chicago Press; 2004:252-272.
34. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]

XSL•FO
RenderX

35. Charles ST, Almeida DM. Daily reports of symptoms and negative affect: Not all symptoms are the same. Psychol Health 2007 Feb 01;21(1):1-17. [doi: 10.1080/14768320500129239]
36. Raudenbush SW, Spybrook J, Congdon R, Liu X, Martinez A, Bloom H, et al. Optimal Design Plus Empirical Evidence (Version 3.0) [Software]. 2011. URL: http://wtgrantfoundation.org/resource/optimal-design-with-empirical-information-od [accessed 2020-02-09]
37. Research Randomizer. URL: https://www.randomizer.org/ [accessed 2021-07-28]
38. Beauchamp M, Hulteen R, Ruissen G, Liu Y, Rhodes R, Puterman E. The Seniors COvid-19 Pandemic and Exercise (SCOPE) study: A randomized controlled trial (SCOPE: Transparent-Changes). 2020 Jul 27. URL: https://osf.io/9fdk4 [accessed 2021-05-23]
39. Muthén LK, Muthén BO. Mplus User's Guide (Sixth Edition). Los Angeles, CA: Muthén & Muthén; 2007.
40. Hu L, Bentler P. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychol Methods 1998;3(4):424-453. [doi: 10.1037/1082-989X.3.4.424]
41. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
42. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53 [FREE Full text] [doi: 10.1037/a0014699] [Medline: 19271847]
43. Feingold A. A regression framework for effect size assessments in longitudinal modeling of group differences. Rev Gen Psychol 2013 Mar;17(1):111-121. [doi: 10.1037/a0030048] [Medline: 23956615]
44. Feingold A. New approaches for estimation of effect sizes and their confidence intervals for treatment effects from randomized controlled trials. Quant Method Psychol 2019;15(2):96-111 [FREE Full text] [doi: 10.20982/tqmp.15.2.p096] [Medline: 32775313]
45. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological distress and loneliness reported by US adults in 2018 and April 2020. JAMA 2020 Jul 07;324(1):93-94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
46. Xie B, Charness N, Fingerman K, Kaye J, Kim MT, Khurshid A. When going digital becomes a necessity: Ensuring older adults' needs for information, services, and social inclusion during COVID-19. J Aging Soc Policy 2020;32(4-5):460-470. [doi: 10.1080/08959420.2020.1771237] [Medline: 32507061]
47. Davidson J, Schimmele C. Evolving Internet Use Among Canadian Seniors. Ottawa, ON: Statistics Canada; 2019 Jul 10. URL: https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2019015-eng.htm [accessed 2020-06-16]

Abbreviations

CES-D: Center for Epidemiologic Studies Depression Scale
CFI: comparative fit index
ePARmed-X+: Electronic Physical Activity Readiness Medical Examination
ES: effect size
FIPPA: Freedom of Information and Protection of Privacy Act
L-CAT: Stanford Leisure-Time Categorical Activity Item
LGM: latent growth model
PARQ+: Physical Activity Readiness Questionnaire for Everyone
RMSEA: root mean square error of approximation
SCOPE: Seniors COVID-19 Pandemic and Exercise
SRMR: standardized root mean square residual
WLC: waitlist control

Edited by C Basch; submitted 25.05.21; peer-reviewed by E Dove, R Menhas; comments to author 17.06.21; revised version received 22.06.21; accepted 14.07.21; published 30.07.21.

Please cite as:
Beauchamp MR, Hulteen RM, Ruissen GR, Liu Y, Rhodes RE, Wierts CM, Waldhauser KJ, Harden SH, Puterman E
Online-Delivered Group and Personal Exercise Programs to Support Low Active Older Adults' Mental Health During the COVID-19 Pandemic: Randomized Controlled Trial
J Med Internet Res 2021;23(7):e30709
URL: https://www.jmir.org/2021/7/e30709
doi:10.2196/30709
PMID:34328433

XSL•FO
RenderX

©Mark R Beauchamp, Ryan M Hulteen, Geralyn R Ruissen, Yan Liu, Ryan E Rhodes, Colin M Wierts, Katrina J Waldhauser, Samantha H Harden, Eli Puterman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stakeholder Perspectives on Barriers and Facilitators for the Adoption of Virtual Clinical Trials: Qualitative Study

Romée Melanie Helena Coert[1,2*], BSc; James Kenneth Timmis[2*], BSc, MSc; André Boorsma[1*], BSc, MSc, PhD; Wilrike J Pasman[1*], BSc, MSc, PhD

[1]Department of Microbiology and Systems Biology, Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek, Zeist, Netherlands
[2]Faculty of Science, Athena Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[*]all authors contributed equally

Corresponding Author:
Romée Melanie Helena Coert, BSc
Faculty of Science
Athena Institute
Vrije Universiteit Amsterdam
De Boelelaan 1105
Amsterdam, 1081 HV
Netherlands
Phone: 31 652688579
Email: romeecoert@hotmail.com

Abstract

Background: Conventional clinical trials are essential for generating high-quality evidence by measuring the efficacy of interventions in rigorously controlled clinical environments. However, their execution can be expensive and time-consuming. In addition, clinical trials face several logistical challenges regarding the identification, recruitment, and retention of participants; consistent data collection during trials; and adequate patient follow-up. This might lead to inefficient resource utilization. In order to partially address the current problems with conventional clinical trials, there exists the need for innovations. One such innovation is the virtual clinical trial (VCT). VCTs allow for the collection and integration of diverse data from multiple information sources, such as electronic health records, clinical and demographic data, patient-reported outcomes, anthropometric and activity measurements, and data collected by digital biomarkers or (small) samples that participants can collect themselves. Although VCTs have the potential to provide substantial value to clinical research and patients because they can lower clinical trial costs, increase the volume of data collected from patients' daily environment, and reduce the burden of patient participation, so far VCT adoption is not commonplace.

Objective: This paper aims to better understand the barriers and facilitators to VCT adoption by determining the factors that influence individuals' considerations regarding VCTs from the perspective of various stakeholders.

Methods: Based on online semistructured interviews, a qualitative study was conducted with pharmaceutical companies, food and health organizations, and an applied research organization in Europe. Data were thematically analyzed using Rogers' diffusion of innovation theory.

Results: A total of 16 individuals with interest and experience in VCTs were interviewed, including persons from pharmaceutical companies (n=6), food and health organizations (n=4), and a research organization (n=6). Key barriers included a potentially low degree of acceptance by regulatory authorities, technical issues (standardization, validation, and data storage), compliance and adherence, and lack of knowledge or comprehension regarding the opportunities VCTs have to offer. Involvement of regulators in development processes, stakeholder exposure to the results of pilot studies, and clear and simple instructions and assistance for patients were considered key facilitators.

Conclusions: Collaboration among all stakeholders in VCT development is crucial to increase knowledge and awareness. Organizations should invest in accurate data collection technologies, and compliance of patients in VCTs needs to be ensured. Multicriteria decision analysis can help determine if a VCT is a preferred option by stakeholders. The findings of this study can be a good starting point to accelerate the development and widespread implementation of VCTs.

(J Med Internet Res 2021;23(7):e26813) doi:10.2196/26813

XSL•FO
RenderX

KEYWORDS

virtual clinical trials; decentralized clinical trials; adoption; do-it-yourself; wearables; diffusion of innovation theory; clinical trials; digital health; virtual health

Introduction

Developers of health interventions (eg, drugs, medical devices, diets, or procedures) have to demonstrate via clinical evidence that their technologies do no, or minimal, harm to patients and improve treatment outcomes [1,2]. Rigorous clinical research, and clinical trials specifically, are necessary to demonstrate sufficient efficacy and safety profiles in order for regulators to grant marketing authorizations and, in turn, for patients to benefit from the introduction of better treatments. However, the execution of a clinical trial is expensive and time-consuming. In addition, clinical trials face several logistical challenges that might lead to inefficient resource utilization [3,4]. These major challenges primarily pertain to the identification, recruitment, and retention of participants; attainment of informed consent; consistent data collection during trials; and adequate patient follow-up [4-6]. Moreover, several demographic groups such as ethnic minorities and elderly patients are underrepresented in clinical trials and therefore the generalizability of trial results might not be immediately clear [4,7,8]. In addition, real-life settings might sometimes be more suitable to comprehensively or appropriately assess the benefits (and risks) of health interventions. Consequently, the ecological validity—capturing data in a real-life setting that reflects the circumstances where the intervention will ultimately occur and across diverse populations—of conventional clinical trials might be limited [6,9]. In order to partially address the current problems with conventional clinical trials, there is a need for innovations in the clinical trial process.

A novel clinical trial concept is the digital clinical trial or virtual clinical trial (VCT), also known as a decentralized, do-it-yourself, remote, siteless, or innovative health trial [4,10,11]. VCTs can be defined as "trials executed through telemedicine, mobile/local healthcare providers and/or mobile technologies" [12,13]. VCTs allow for the collection and integration of diverse data from multiple information streams, such as electronic health records, clinical and demographic data, patient-reported outcomes, anthropometric and activity measurements, and data collected by digital biomarkers or (small) samples that participants can collect themselves (eg, blood droplet, saliva, or fecal sample sent for analysis to a lab) [4]. They could be used, for example, in dermatology research where skin diseases can be evaluated remotely [12]. In a study from Singer and colleagues [14], 69 participants could easily take photos of their skin after treatment and send them to a physician for evaluation. Furthermore, VCTs might also be useful in nutrition research where participants follow a diet and perform exercises or physical activity at home guided or monitored by health technologies such as mobile apps, wearables, online data collection, or web-based tests [15,16]. Such digital diagnostics and tools have to meet high validity and reliability quality standards that might be difficult to attain before they can be applied in these new types of trials [4]. In addition, data protection could be an issue since large amounts of sensitive health information might be transferred [12]. Nevertheless, the proper implementation of VCTs can reduce trial costs per participant by up to 50% (compared to conventional clinical trials) because many relevant tests can be performed and evaluated without the need for patients to visit specific sites [17-19]. This also means that less time from health care professionals is needed for data collection. Additionally, a substantially increased volume of data can be collected from patients' daily environment and, thereby, potentially provide early estimation of intervention effectiveness [18,19]. Moreover, digital health technologies, such as wearables, can provide continuous monitoring of trial participants to rapidly identify adverse events [4]. Because, for example, recruitment and enrolment of patients, and (long-term) follow-up can be done remotely, the burden of patient participation is reduced, and participant diversity and retention can be significantly improved [10,20]. Importantly, VCTs can also deliver a more patient-centered approach and engage patients in lifestyle and clinical research, and thereby improve their overall health literacy [21]. Lastly, the recent COVID-19 pandemic further highlights the advantages of VCTs for various stakeholders: although clinical trial organizations had to put their traditional site trials on hold, virtual visits and data collection were still partially possible online [22,23]. As a consequence, Izmailova and colleagues [23] developed a decision tree for migration from clinic to remote activities.

Although VCTs have the potential to provide huge value to clinical research and, by extension, patients, VCT adoption is not commonplace as of yet [4]. This is surprising, as failing to explore all options of health intervention advancements runs the risk of missing opportunities to preserve and promote patients' health and improve overall resource distribution in health care. Research into the perspectives of stakeholders such as research organizations, pharmaceutical organizations, and food and health organizations is needed. Previous research from the Clinical Trials Transformation Initiative (CTTI) was focused on the perspectives of sponsors to identify legal and regulatory challenges of VCTs, but research organizations and food and health organizations were not included [13]. Therefore, the aim of this study was to better understand barriers and facilitators to the adoption of VCTs by determining the factors that influence individuals' considerations regarding VCTs from the viewpoints of pharmaceutical organizations, food companies, and a research organization, by taking a research organization's perspective. The research organization Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (TNO; Netherlands Organisation for Applied Scientific Research) is an independent, government-funded applied research organization that aims to nurture innovation by closing the gap between research and industry. TNO connects companies and knowledge in order to create innovations that sustainably strengthen the competitiveness of the high tech sector and, in turn, improve the well-being of society [24]. This study was a

XSL•FO
RenderX

first pilot investigation to examine the readiness of companies and organizations to adopt VCTs.

Methods

Study Overview

We conducted a qualitative study on the perspectives of stakeholders about the adoption of VCTs. The Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist for reporting qualitative research was followed (Multimedia Appendix 1) [25]. Approval for the qualitative study was granted by the Internal Review Board of TNO in April 2020 (reference number: 2020-034).

Theoretical Considerations

To base our investigation on a solid theoretical framework, we utilized Rogers' diffusion of innovation theory. Rogers' framework aims to explain how and why a new innovation or technology is, or is not, adopted, where adoption can be defined as "the full use of an innovation [when the innovation] is the best course of action available" [26]. A VCT can be considered a process innovation, which is defined as the implementation of a new or improved delivery method, including substantial changes in techniques, equipment, and/or software [27]. Zhang and colleagues [28] reported their investigation of the factors leading to a successful eHealth process innovation. They used Rogers' theory to analyze patient acceptance and identify reasons for the utilization of eHealth innovations [28]. Several other recent studies have also used Rogers' framework to better understand health technology adoption [28,29]. Therefore, Rogers' theory was considered to be an appropriate framework for this study. Multimedia Appendix 2 [30-33] provides more details about Rogers' theory and concepts, including the five stages of the innovation-decision process.

Data Sources and Sampling Strategy

A purposive sampling strategy was used to select invitees [25]. Three stakeholder groups were approached via email: invitees were recruited from TNO (within which people working in the health division were approached), pharmaceutical organizations, and food and health organizations. Within TNO, the Microbiology and Systems Biology (MSB) department supports the research and development activities of companies in the agriculture and food, health, personal care, chemistry, biotechnology, and pharmaceutical sectors. The department collaborates with pharmaceutical, food, and health organizations as well as other companies to set up clinical trials [24]. Within the pharmaceutical industry, a distinction between sponsors and contract research organizations (CROs) can be made. Food and health companies can be defined as organizations that aim to improve health by healthy foods. These organizations study nutrition by conducting nutrition trials.

Twenty potential participants received an interview invitation. Due to the COVID-19 pandemic, all interviews were held by telephone or Skype. Invitees were enrolled in the study if they met the following inclusion criteria: speak Dutch or English fluently; work for one of the three stakeholder groups (research organization, pharmaceutical industry, or a food company) in a key position of clinical trial execution; have experience and/or interest in using VCTs; and are willing and able to participate and sign an online informed consent form.

Data Collection and Instrument

The semistructured interviews were carried out by author RMHC. The interview guide consisted of themes and questions based on the five innovation stages as suggested by Rogers, namely: knowledge, persuasion, decision, implementation, and confirmation (Multimedia Appendix 3). The interview guide was discussed with and approved by all authors. Furthermore, the interview was piloted with 2 independent researchers at TNO [25]. Once the invitees had given consent, the interviews were conducted and audio recorded. Field notes were made during and directly after the interviews.

Data Analysis

The audio recordings were transcribed immediately after the interview, new concepts were identified, and the degree of data saturation was measured [34-37]. In addition, to confirm the accuracy of content and key messages, participants were asked to review a brief summary of the interview and a couple of quotes, and comment if necessary [35,37]. The recordings were deleted directly after transcription, and the transcripts were stored as Microsoft Word files (Microsoft Corp) on a secured laptop per proper data and privacy protection measures. The data were analyzed by using the thematic analysis method suggested by Clarke and Braun [38], which consists of the following 6 steps: familiarization with the data, initial code generation, theme identification, theme review, theme definition and naming, and, finally, report production [38]. The data obtained from the interviews were coded by RMHC and merged into a code book. A 5% check (1 interview transcript and coding) was done by another researcher to ensure that coding was performed consistently. The data from the interviews were descriptively analyzed with Microsoft Excel 2016 (Microsoft Corp).

Results

Participant Characteristics

From the 20 potential participants who were approached, 4 individuals were excluded from this study; 2 did not respond to the invitation, and 2 were not experienced enough in VCTs and thus did not met the inclusion criteria. In total, 16 individuals were interviewed between April and June 2020. Six employees from a research organization, 6 pharmaceutical employees, and 4 individuals from food companies were enrolled in the study. Among the pharmaceutical participants, one was an employee of a CRO. This person was grouped into the pharmaceutical cohort since the results of the interviews between the CRO and pharmaceutical companies were much aligned. All interviewees from the research organization lived in The Netherlands, while 2 pharmaceutical interviewees and 2 interviewees from food companies were based in Ireland, Denmark, France, and Germany (Table 1). In Table 2, the type of organizations where interviewees were employed is shown.

XSL•FO
RenderX

Table 1. Participant characteristics (N=16).

Characteristic	Participants, n (%)
Organization	
Research organization	6 (37.5)
Junior	0 (0)
Senior[a]	6 (37.5)
Pharmaceutical companies	6 (37.5)
Junior	2 (12.5)
Senior	4 (25.0)
Food companies	4 (25.0)
Junior	2 (12.5)
Senior	2 (12.5)
Country	
The Netherlands	12 (75.0)
Other parts of Europe (Ireland, Denmark, France, and Germany)	4 (25.0)
Gender	
Male	13 (81.25)
Female	3 (18.75)

[a]A senior has more than 5 years of experience in his or her job.

Table 2. Type of organizations where interviewees were employed (N=11).

Organization type	Organizations included, n (%)
Research organization	1 (9.1)
Pharmaceutical companies	6 (54.5)
Food companies	4 (36.4)

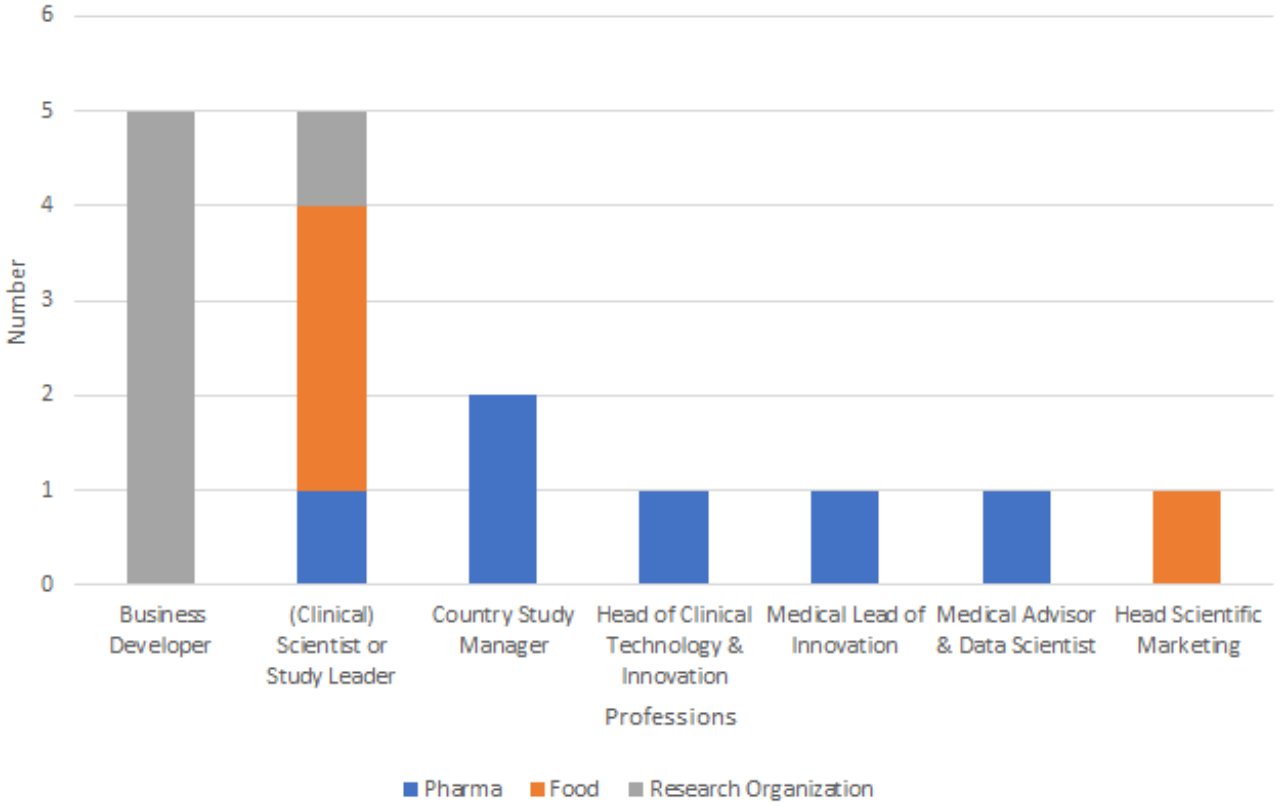
Of the 6 individuals affiliated with the research organization, 5 were business developers. Business developers were included since they are the initial point of contact with business partners and discuss different options for conducting studies. They have ideas about how VCTs could be interesting for their clients and whether the research organization should be involved in VCTs. The remaining individual was a senior scientist conducting clinical research in collaboration with pharmaceutical companies and universities.

The 6 participants from the pharmaceutical companies had various professional backgrounds. Two were country study managers leading a team that conducts clinical trials. Moreover, they are responsible for the quality of the research. Two participants worked within an innovation department. One was the medical lead of innovation, and the other worked as the head of clinical technology and innovation. The medical lead of innovation was responsible for several (clinical trial) innovation projects within the company. The head of clinical technology and innovation was closely involved with conducting VCTs. Another interviewee was a clinical scientist, who was involved in conducting clinical trials. The last participant was a medical advisor and data scientist, and was responsible for coordinating clinical research and analyzing clinical trial data.

Three of the 4 interviewees from the food companies were (clinical) scientists and/or study leaders. They were responsible for conducting or leading clinical trials. The remaining participant was head of scientific marketing within a food company, who conceptualized, planned, and implemented clinical research. Figure 1 displays the professional backgrounds of the participants.

XSL•FO
RenderX

Figure 1. Overview of participants' professions.



A total of 12 interviews were conducted by telephone and 4 by Skype (with video). Interviews lasted approximately 45 minutes. After the 15th interview, data saturation was reached. Data collection was continued for one more interview to confirm that no new concepts were mentioned.

Overview of Findings

Table 3 illustrates current barriers and facilitators, according to stakeholder perspectives.

Table 3. Barriers and facilitators for the adoption of virtual clinical trials. Counts indicate how many interviewees agreed on certain barriers and facilitators.

Barriers and facilitators	Research organization (n=6), n (%)	Pharmaceutical companies (n=6), n (%)	Food companies (n=4), n (%)	Total across all organization types, n (%)
Barriers				
Acceptance by regulatory authorities	3 (18.75)	3 (18.75)	2 (12.5)	8 (50.0)
Technical issues (standardization, validation, and data storage)[a]	4 (25.0)	4 (25.0)	1 (6.25)	9 (56.3)
Compliance and adherence[a]	5 (31.25)	6 (37.5)	3 (18.75)	14 (87.5)
Lack of knowledge or understanding	2 (12.5)	1 (6.25)	3 (18.75)	6 (37.5)
Facilitators				
Involving regulators in the development process	3 (18.75)	3 (18.75)	1 (6.25)	7 (43.8)
Exposure to the results of pilot studies[a]	5 (31.25)	2 (12.5)	3 (18.75)	10 (62.5)
Clear instructions and assistance for patients[a]	4 (25.0)	6 (37.5)	3 (18.75)	13 (81.3)

[a]Barriers and facilitators that many stakeholders agreed on (ie, more than half [>8] of the interviewees mentioned it).

Barriers to VCT Adoption

Acceptance by Regulatory Authorities

The concern of many participants was that regulatory authorities such as the Food and Drug Administration (FDA) and European Medicines Agency are not ready to accept VCTs because the procedures to evaluate VCTs are currently not standardized or well established. One person from a pharmaceutical company mentioned that, although some guidelines were available, they were vague and abstract, and this made it highly challenging to estimate if a VCT would be approved. For the pharmaceutical industry, this was important as evidence would have to be

XSL•FO
RenderX

obtained by approved methods; otherwise pharmaceutical products cannot receive approval.

Technical Issues: Standardization, Validation, and Data Storage

Participants mentioned several technical issues regarding the execution of VCTs. One of them was the lack of standardization. Participants expressed that the development of smart devices is fast-paced and, in addition, the algorithms of these devices are not always known, which could lead to these devices being considered a "black box." Therefore, respondents explained that the data collected today would become outdated in a year's time as new algorithms and smart devices become available. In addition, investigators are mandated to store study data for 15 years, but if the technology becomes outdated and cannot be updated, data storage (and access) can be problematic. Another issue mentioned was accessibility of data through an app or wearable developer, whereby the researcher has no access to the propriety data. Respondents were also concerned about the accuracy of wearables; although many wearables are currently available, most have not been validated. If a wearable did not have formal validation, the data it generated could not be used as clinical evidence.

Compliance or Adherence

Respondents from all organization types mentioned that even in normal clinical trials, compliance (the degree to which a patient correctly follows an intervention) was an issue. In a VCT, they believed that this could become an even bigger issue, as in-person contact with participants is limited. For example, the drug would have to be taken, or the wearable used, by the trial participant only, which is challenging to monitor in a decentralized setting:

> *If you conduct it [a trial] in a clinical setting, there is someone [a study nurse/researcher] standing next to the participant, so he or she checks if everything is done correctly. But if you do it [ie, a trial] at home ... you [a researcher] don't see it. Or, for example, you [a researcher] track a mobile phone, it can also remain on the table, in another room.* [P3]

Lack of Knowledge or Understanding

The interviews revealed that some organizations, mainly food companies and the research organization, had a lack of knowledge regarding the opportunities offered by VCTs. Parts of clinical trials could be done remotely, but food companies and some interviewees from the research organization were not familiar with conducting clinical trials remotely and/or the development of remote technologies. For example, the concrete utility of, and how to develop, VCTs remained unclear for food companies, as they were not aware of other organizations that utilize VCTs and did not know how to obtain relevant information. Therefore, they were more reluctant to explore VCTs. This indicates that collaboration between stakeholders, especially between food companies and the pharmaceutical industry, were not effective. Pharmaceutical companies, on the other hand, were aware of the opportunities offered by VCTs and knew that others were exploring this field and developing new platforms. Participants from all stakeholder groups mentioned that some patient groups might face various issues related to VCTs. In particular, elderly patients may not have the experience that is required, such as knowing how to use a smartwatch or the internet, for this study type, even though it would be highly preferable to include this patient group in trials. Consequently, a relevant part of the population might not be included in VCTs, which could lead to biased data and thus imprecise estimates of intervention effectiveness.

Facilitators of VCT Adoption

Involving Regulators and Other Stakeholders in Development Processes

According to participants, the acceptance of VCTs could be facilitated by involving regulators early in the development phase of a VCT. Furthermore, it was also important to collaborate with other stakeholders during this phase such as other pharmaceutical companies, research organizations, and food companies, in addition to also incorporating the clinic (doctors and patients) to investigate their needs:

> *About 15 years ago, pharmaceutical companies were loose [isolated] strongholds. But now you see more and more that the companies are working together [open innovation]. And I think collaboration is the future. It cannot be otherwise because the development of new innovative medicines is very expensive. You just have to do it together.* [P1]

Exposure to the Results of Pilot Studies

According to the pharmaceutical participants, many of them were currently doing validation or pilot studies with wearables, online platforms (eg, for trial recruitment), or mobile apps. Conducting such pilot studies and publishing the results would support other organizations in understanding VCTs and how to use them. It might also convince regulators that the collected data were reliable and added value to new health interventions.

Clear Instructions and Assistance for Patients

All three stakeholder groups indicated that patients in a VCT needed additional assistance from the researcher or study nurse with, for instance, the proper use of a wearable. One participant (pharmaceutical company) indicated that guidance was needed because the tool (eg, wearable) itself could be useless without user input. Specific groups such as elderly trial participants need proper instructions and guidance (eg, via a helpdesk).

The Decision to Adopt VCTs

A point mentioned by participants across all organizations was that adoption depended on the type of study to be performed. In principle, if a research question was suitable, for instance, if a researcher was interested in studying lifestyle changes measured through a wearable or collected via an online questionnaire at home, a VCT could provide a significantly richer data set (including real-life data) when compared to a conventional clinical trial. On a more practical note, for many respondents, the decision to adopt VCTs also depended on the costs of the trials. If VCTs could in effect reduce costs, some respondents would be inclined to conduct a VCT rather than a conventional clinical trial.

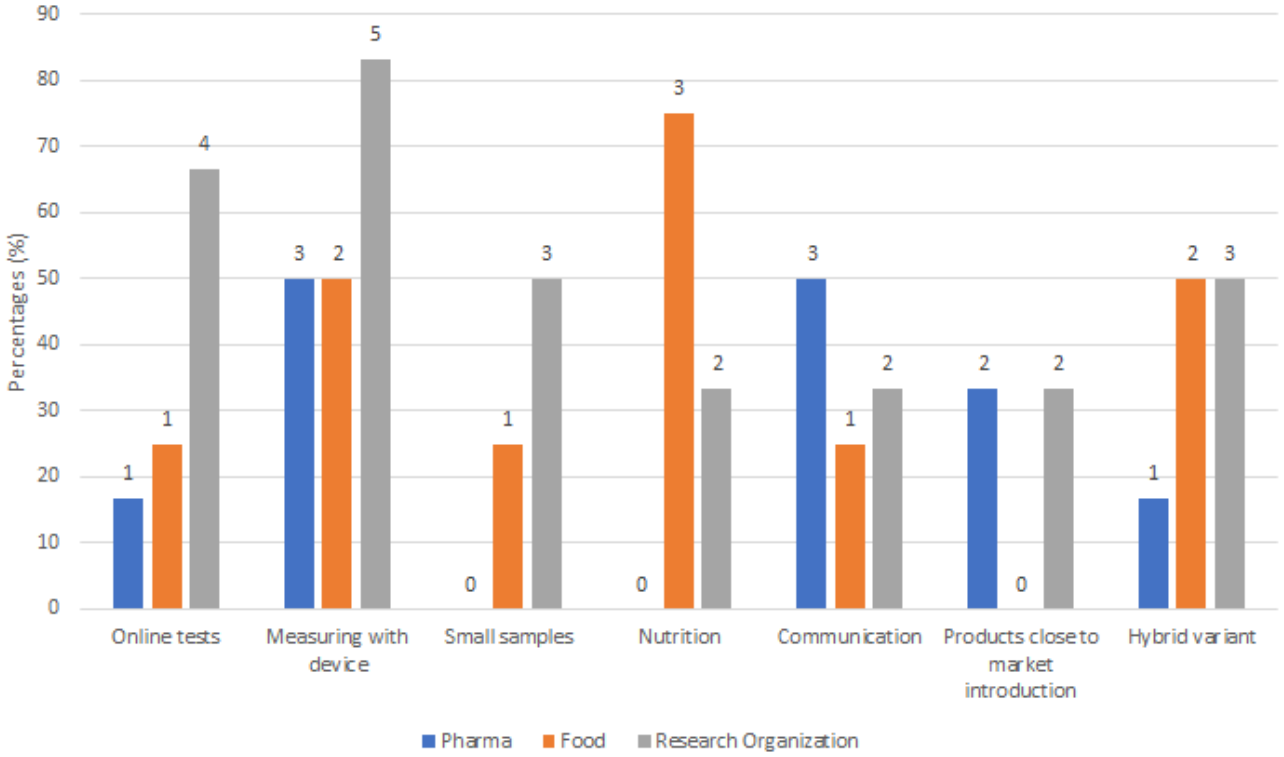
XSL•FO
RenderX

Options to Apply VCTs

The three stakeholder groups made several suggestions on what types of tools could be used and how VCTs could be implemented. Contemporary smart devices or wearables could be used to measure health parameters (eg, heart rate, temperature, oxygen saturation, respiration rate, blood pressure, blood glucose level) and monitor patients while the investigational intervention or drug is studied. Next to noninvasive measurements, the collection of small samples, such as drawing blood by a finger prick, is also possible. By extension, multiple blood droplet analysis devices exist that enable consumers to collect data (eg, blood cholesterol, glucose, or ferritin) themselves and at home. Furthermore, participants mentioned the possibility of conducting online tests consisting of questionnaires or cognition assessments that could be used in, for example, nutrition studies, where participants consume a daily dietary supplement. The effect of the supplement on satiety or gastrointestinal function could be monitored easily online and does not require clinical assistance.

Communication between patients and researchers/doctors could be executed virtually (ie, online). Visits to the clinic, which usually would be planned in conventional clinical trials, could be done remotely, especially for follow-up visits once the patient is considered disease-free. Respondents from the pharmaceutical companies and the research organization explained that a VCT could be implemented when a product or service was close to market introduction. Such trials might investigate how the product or service works in the real-life environment of patients. The hybrid variant was mentioned as the most promising form of a VCT. This means that such trials would have virtual and conventional clinical trial components. For instance, conventional clinical trials could have a hybrid locality approach by collecting some data outside of the central research facilities (eg, plasma and tissue samples are collected by affiliated labs and sent to central locations for testing). Figure 2 shows agreement among the interviewees about potential VCT options.

Figure 2. Options related to virtual clinical trials (VCTs) that interviewees agreed on. The counts and percentages provide an indication of how many participants suggested VCTs for a specific study type.



Discussion

Principal Results

In this study, participants acknowledged the innovative value of VCTs and agreed that the development of VCTs should be further explored and promoted. However, various aspects have to be improved before VCTs can be widely adopted, such as improving effective collaboration between all stakeholders, clinically validating smart devices and wearables for data collection, and ensuring patient compliance. The data obtained from our interviewees demonstrate that pharmaceutical companies and the research organization in particular are aware of the benefits and disadvantages of VCTs. However, the applicability of VCTs in research is not yet well known. Therefore, informing stakeholders of the advantages of VCTs is an important factor to speed up their consideration and adoption. This is relevant especially now during the COVID-19 pandemic, due to which many conventional randomized controlled trials have been halted; here, VCTs can make a substantial contribution to the continuation of certain types of crucial clinical research [22].

XSL•FO
RenderX

Comparison With Prior Work

Collaboration

All participants agreed on the importance of enhancing collaboration among stakeholders and, in particular, with regulatory authorities. The acceptance of VCTs by regulators was perceived as a barrier, which is in line with previous research. Polhemus et al [39] studied the adoption of patient-facing technologies in clinical trials from an industry perspective, and one of the main barriers they identified were regulatory challenges. This was due to a lack of specific guidance (no or few regulations), geographic variability of guidance, and internal perceptions and misperceptions [39]. In addition, the CTTI stated that telemedicine was not widely used in the design and conduct of clinical trials because of legal and regulatory considerations [40]. Due to the COVID-19 pandemic, medical centers and CROs are required to set up online communication and information repositories, resulting in the release of new VCT guidelines by the FDA [22,41]. When designing VCTs, it is important to take these guidelines into consideration. The data infrastructure, processing, analysis, and interpretation of a VCT are different compared to a conventional clinical trial. As also proposed by the CTTI, there is a need for collaboration across various experts such as CROs, pharmaceutical organizations, research organizations, and food and health organizations on one hand, and the companies developing new technologies (eg, wearables) on the other [13,42,43].

Validation of Data Collection Tools

All participants emphasized that the accuracy of the data collected from, for example, wearables; do-it-yourself blood analysis of glucose, cholesterol, hemoglobin, etc; or blood pressure devices in VCTs, is crucial. Currently, there exist many consumer-grade devices that promise to improve health and wellness without scientific evidence substantiating such claims; hence, there is an urgent need for clinically validated devices and wearables [9,43]. The quality assurance (and measurement accuracy) of technologies has been questioned before and several challenges were identified by previous studies [39,44]. The findings of Abdolkhani et al [44] indicated that technical and policy standards need to be developed to guarantee the quality of data generated from wearables. Besides the need for such standards, it is important to validate the wearables by "fit-for-purpose validation" [45]. In such trials, it should become clear, within a very short time frame, if the technology is suitable to accurately measure clinically meaningful endpoints. According to Goldsack et al [45], the evaluation of a wearable device, or a biometric monitoring device, should consist of a three-component framework, consisting of verification, analytical validation, and clinical validation [45]. This has to be completed before technologies can be utilized in VCTs.

Compliance

As described by our interviewees, investigators must carefully consider how to ensure a high degree of compliance in VCTs. Moreover, they explained that proper instructions are crucial as patients have to fully understand what is expected of them as active partners in a trial and need to know how to obtain information. A clear and simple protocol is therefore required [46]. According to previous research, participants who understand the expectations of a clinical trial are more willing and able to comply [47]. Furthermore, some participant groups, such as the elderly can be less literate regarding electronic and smart devices [48,49]. In order to overcome this challenge, interviewees suggested that these participant groups receive more assistance during VCTs. The technology used in a VCT needs to be user-friendly, and the trial itself needs to be as simple as possible [48]. Therefore, it has been recommended by the CTTI that participants be engaged in technology selection. Only technologies that are easy to learn, simple, convenient to use, and physically comfortable should be included in a VCT [46,50].

Implications for Practice and Policy

In order to improve the adoption of VCTs, it is important that all stakeholders collaborate with each other [51]. One option for enhancing collaboration among all stakeholders and exploring if a VCT is a preferred option is an multicriteria decision analysis (MCDA) [52]. Although we have not yet implemented this method, an MCDA might be a good option for researchers who are considering VCTs as a data collection tool. An MCDA is a set of techniques that can help decision-makers take into account and integrate multidimensional data (eg, attributes of benefit) and rank different decision alternatives [52]. For example, the preferences of different criteria or parameters relevant to clinical trials could be elicited a priori and subsequently "preloaded" as preference profiles into dedicated MCDA software. This would potentially result in a more transparent and rapid decision process, and more effective support for decision-makers in determining if a VCT, or some of its components, would be beneficial for their trial. Although an MCDA gives a transparent methodology to compare different decision alternatives, it is an extensive technique that might not be applicable to every context; one setting might be more appropriate for an MCDA than another. For instance, it could be more easily implemented in larger organizations with, in relative terms, larger budgets than in smaller organizations. Next to the MCDA, stakeholders should invest in the use of validated diagnostics to obtain the most reliable results in VCTs [45]. The adherence of patients can be improved by properly devising tools that can motivate them (eg, in a playful way, that is, gamification) or via online monitoring. Because a VCT is a new form of clinical trial with a digital approach, additional assistance might be necessary for some participants. For this reason, we suggest the installation of helpdesks to support trial participants by answering their questions and providing them with support 24/7.

Strengths, Limitations, and Future Research

Underpinned by a well-known theoretical framework of innovation, our qualitative research approach allowed us to elicit and characterize the broad experiences of interviewees with the VCT adoption process [35]. Furthermore, we made use of the COREQ checklist and an interview guide, which improved the validity and standardization of the interviews, while allowing improvised follow-up questions based on our interviewees' responses [25,35]. Whereas most studies

XSL•FO
RenderX

concerning VCTs are conducted in the United States, our research focused on Europe, which is unique.

The conduct of this research in the European Union could be framed as a strength but also as a limitation. Data sampling was restricted to Europe, mostly to The Netherlands. Since the United States is more advanced in conducting VCTs, follow-up studies should also include this country in their data sample. Furthermore, this pilot study only included pharmaceutical companies (including a CRO), food companies, and a single research organization. Health care professionals, patients or participants, regulatory authorities, CROs, and payers are also important stakeholders in VCTs. According to previous studies, these stakeholders should also be included in the process of clinical trial development [42,53]. Therefore, future research should focus on these stakeholders, and their preferences, value points, and perspectives on VCTs. Lastly, inclusion of only one research organization introduced selection bias, leading to a less representative study sample.

Conclusions

This study used a qualitative research approach to identify the barriers and facilitators behind the adoption of VCTs and explored how this process can be improved. Collaboration among all stakeholders in VCT development is essential to increase knowledge and awareness. Organizations should invest in accurate data collection technologies, and compliance of patients in VCTs needs to be ensured. Furthermore, we suggest conducting an MCDA to explore whether a VCT is a preferred option by stakeholders; this can considerably enhance the decision-making process. The findings of this study can be a good vantage point to accelerate the development and widespread implementation of VCTs.

Acknowledgments

The authors thank all the participants of the study for their time and effort.

Authors' Contributions

RMHC conducted the interviews and analyzed the data. JKT, WJP, and AB were involved in all stages of the research project by providing feedback and instructions to RMHC. RMHC wrote the first draft of the paper. All authors participated in reviewing and improving all versions of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ checklist.
[DOCX File , 177 KB - jmir_v23i7e26813_app1.docx]

Multimedia Appendix 2
Rogers' diffusion of innovation theory.
[DOCX File , 101 KB - jmir_v23i7e26813_app2.docx]

Multimedia Appendix 3
Interview guide.
[DOCX File , 13 KB - jmir_v23i7e26813_app3.docx]

References

1. Learn About Clinical Studies. US National Library of Medicine, ClinicalTrials.gov. 2019. URL: https://clinicaltrials.gov/ct2/about-studies/learn [accessed 2020-03-11]
2. Friedman LM, Furberg CD, Demets DL, Reboussin DM, Granger CB. Chapter 1: Introduction to Clinical Trials. In: Fundamentals of Clinical Trials. Cham, Switzerland: Springer; 2015:1-14 URL: https://tinyurl.com/3d6cmsfb
3. Brucher C, Duprel C, Georges A, Kilders Y. Digitalization of clinical trials: How new technologies enhance the future of health care. Inside Magazine. 2018 Nov. URL: https://www2.deloitte.com/lu/en/pages/life-sciences-and-healthcare/articles/digitalization-clinical-trials.html [accessed 2020-02-12]
4. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3(1):101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
5. To Get Better Results, Talk To The Real Audience. Clinical Leader. 2018. URL: https://www.clinicalleader.com/doc/clinicaltrials-gov-to-get-better-results-talk-to-the-real-audience-0001 [accessed 2020-08-14]
6. Pasman WJ, Boessen R, Donner Y, Clabbers N, Boorsma A. Effect of Caffeine on Attention and Alertness Measured in a Home-Setting, Using Web-Based Cognition Tests. JMIR Res Protoc 2017 Sep 07;6(9):e169-e169 [FREE Full text] [doi: 10.2196/resprot.6727] [Medline: 28882811]

XSL•FO
RenderX

7. Nazha B, Mishra M, Pentz R, Owonikoko TK. Enrollment of Racial Minorities in Clinical Trials: Old Problem Assumes New Urgency in the Age of Immunotherapy. American Society of Clinical Oncology Educational Book 2019 May(39):3-10. [doi: 10.1200/edbk_100021]
8. Farkouh ME, Fuster V. Time to welcome the elderly into clinical trials. Nat Clin Pract Cardiovasc Med 2008 Nov;5(11):673-673. [doi: 10.1038/ncpcardio1387] [Medline: 18949002]
9. Kruizinga MD, Stuurman FE, Groeneveld GJ, Cohen AF. The Future of Clinical Trial Design: The Transition from Hard Endpoints to Value-Based Endpoints. Handb Exp Pharmacol 2019;260:371-397. [doi: 10.1007/164_2019_302] [Medline: 31707472]
10. Persky S. A Virtual Home for the Virtual Clinical Trial. J Med Internet Res 2020 Jan 03;22(1):e15582 [FREE Full text] [doi: 10.2196/15582] [Medline: 31899455]
11. Steinhubl SR, Wolff-Hughes DL, Nilsen W, Iturriaga E, Califf RM. Digital clinical trials: creating a vision for the future. NPJ Digit Med 2019 Dec 12;2(1):126 [FREE Full text] [doi: 10.1038/s41746-019-0203-0] [Medline: 31872066]
12. Ali Z, Zibert J, Thomsen S. Virtual Clinical Trials: Perspectives in Dermatology. Dermatology 2020 Mar 3;236(4):375-382 [FREE Full text] [doi: 10.1159/000506418] [Medline: 32126560]
13. Apostolaros M, Babaian D, Corneli A, Forrest A, Hamre G, Hewett J, et al. Legal, Regulatory, and Practical Issues to Consider When Adopting Decentralized Clinical Trials: Recommendations From the Clinical Trials Transformation Initiative. Ther Innov Regul Sci 2020 Jul 9;54(4):779-787 [FREE Full text] [doi: 10.1007/s43441-019-00006-4] [Medline: 32557302]
14. Singer HM, Almazan T, Craft N, David CV, Eells S, Erfe C, et al. Using Network Oriented Research Assistant (NORA) Technology to Compare Digital Photographic With In-Person Assessment of Acne Vulgaris. JAMA Dermatol 2018 Feb 01;154(2):188-190 [FREE Full text] [doi: 10.1001/jamadermatol.2017.5141] [Medline: 29261843]
15. Carrera PM, Dalton AR. Do-it-yourself healthcare: the current landscape, prospects and consequences. Maturitas 2014 Jan;77(1):37-40. [doi: 10.1016/j.maturitas.2013.10.022] [Medline: 24287177]
16. Pasman WJ, Hendriks HF, Minekus MM, de Ligt RA, Scholtes-Timmerman MJ, Clabbers ND, et al. Subjective feelings of appetite of wholegrain breakfasts evaluated under controlled, laboratory and 'at home' conditions. Physiol Behav 2018 Oct 01;194:285-291. [doi: 10.1016/j.physbeh.2018.06.024] [Medline: 29913230]
17. Elenko E, Underwood L, Zohar D. Defining digital medicine. Nat Biotechnol 2015 May;33(5):456-461. [doi: 10.1038/nbt.3222] [Medline: 25965750]
18. Khozin S, Coravos A. Decentralized Trials in the Age of Real-World Evidence and Inclusivity in Clinical Investigations. Clin Pharmacol Ther 2019 Jul 23;106(1):25-27. [doi: 10.1002/cpt.1441] [Medline: 31013350]
19. Coravos A, Goldsack JC, Karlin DR, Nebeker C, Perakslis E, Zimmerman N, et al. Fast Facts: Digital Medicine Measurement. Abingdon, Oxford, UK: Karger Publishers; 2020:1-96.
20. Sommer C, Zuccolin D, Arnera V, Schmitz N, Adolfsson P, Colombo N, et al. Building clinical trials around patients: Evaluation and comparison of decentralized and conventional site models in patients with low back pain. Contemp Clin Trials Commun 2018 Sep;11:120-126 [FREE Full text] [doi: 10.1016/j.conctc.2018.06.008] [Medline: 30094387]
21. Donnelly S, Reginatto B, Kearns O, Mc Carthy M, Byrom B, Muehlhausen W, et al. The Burden of a Remote Trial in a Nursing Home Setting: Qualitative Study. J Med Internet Res 2018 Jun 19;20(6):e220 [FREE Full text] [doi: 10.2196/jmir.9638] [Medline: 29921563]
22. Ledford H. The coronavirus outbreak could make it quicker and easier to trial drugs. Nature 2020 Jun 4;582(7811):172-172. [doi: 10.1038/d41586-020-01524-0] [Medline: 32504019]
23. Izmailova ES, Ellis R, Benko C. Remote Monitoring in Clinical Trials During the COVID-19 Pandemic. Clin Transl Sci 2020 Sep;13(5):838-841 [FREE Full text] [doi: 10.1111/cts.12834] [Medline: 32526077]
24. TNO. 2020. URL: https://www.tno.nl/nl/over-tno/ [accessed 2020-02-19]
25. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
26. Sahin I. Detailed Review of Rogers' Diffusion of Innovations Theory and Educational Technology-Related Studies Based on Rogers' Theory. The Turkish Online Journal of Educational Technology 2006 Apr;5(2):14-23 [FREE Full text]
27. Process Innovation. OECD Glossary of Statistics. 2005 Sep 09. URL: https://stats.oecd.org/glossary/detail.asp?ID=6870 [accessed 2020-11-07]
28. Zhang X, Yu P, Yan J, Ton A M Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
29. Chew F, Grant W, Tote R. Doctors on-line: using diffusion of innovations theory to understand internet use. Fam Med 2004 Oct;36(9):645-650 [FREE Full text] [Medline: 15467942]
30. Rogers E. Diffusion of Innovations. New York, NY: Simon & Schuster; 2003.
31. Rogers E. Diffusion of Innovation Theory (Figure 4). ResearchGate. 1983. URL: https://www.researchgate.net/figure/Diffusion-of-Innovation-Theory-Roger-1983_fig4_286777535 [accessed 2020-02-28]

XSL•FO
RenderX

32. Scott SD, Plotnikoff RC, Karunamuni N, Bize R, Rodgers W. Factors influencing the adoption of an innovation: an examination of the uptake of the Canadian Heart Health Kit (HHK). Implement Sci 2008 Oct 02;3(1):41 [FREE Full text] [doi: 10.1186/1748-5908-3-41] [Medline: 18831766]
33. Kaminski J. Diffusion of Innovation Theory: Theory in Nursing Informatics Column. Can J Nurs Informatics 2011;6(2):1-7 [FREE Full text]
34. Bai X, Luo Z, Wang A, Guan Z, Zhong Z, Sun M, et al. Challenge of parents caring for children or adolescents with early-stage schizophrenia in China: A qualitative study. Perspect Psychiatr Care 2020 Oct;56(4):777-784. [doi: 10.1111/ppc.12492] [Medline: 32147848]
35. Gray DE. Doing Research in the Real World. London, UK: SAGE Publications; 2012.
36. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Mar;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
37. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Polhemus AM, Kadhim H, Barnes S, Zebrowski SE, Simmonds A, Masand SN, et al. Accelerating Adoption of Patient-Facing Technologies in Clinical Trials: A Pharmaceutical Industry Perspective on Opportunities and Challenges. Ther Innov Regul Sci 2019 Jan 31;53(1):8-24. [doi: 10.1177/2168479018801566] [Medline: 30373404]
40. CTTI Recommendations: Decentralized Clinical Trials. CTTI. 2018 Sep. URL: https://www.ctti-clinicaltrials.org/projects/decentralized-clinical-trials [accessed 2020-06-08]
41. FDA Guidance on Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency. FDA. 2020 Mar. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medical-products-during-covid-19-public-health-emergency [accessed 2020-06-02]
42. Huang GD, Bull J, Johnston McKee K, Mahon E, Harper B, Roberts JN, CTTI Recruitment Project Team. Clinical trials recruitment planning: A proposed framework from the Clinical Trials Transformation Initiative. Contemp Clin Trials 2018 Mar;66:74-79 [FREE Full text] [doi: 10.1016/j.cct.2018.01.003] [Medline: 29330082]
43. Izmailova ES, Wagner JA, Perakslis ED. Wearable Devices in Clinical Trials: Hype and Hypothesis. Clin Pharmacol Ther 2018 Jul 02;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
44. Abdolkhani R, Gray K, Borda A, DeSouza R. Quality Assurance of Health Wearables Data: Participatory Workshop on Barriers, Solutions, and Expectations. JMIR Mhealth Uhealth 2020 Jan 22;8(1):e15329 [FREE Full text] [doi: 10.2196/15329] [Medline: 32012090]
45. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
46. Cerreta F, Ritzhaupt A, Metcalfe T, Askin S, Duarte J, Berntgen M, et al. Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 2020 Sep 12;19(9):573-574. [doi: 10.1038/d41573-020-00080-6] [Medline: 32398879]
47. Anderson D. Want to Improve Clinical-Trial Retention and Compliance? Make Convenience a Priority. Deloitte. 2020 Feb. URL: https://www2.deloitte.com/us/en/blog/health-care-blog/2020/improve-clinical-trial.html [accessed 2020-06-11]
48. Novel Clinical Trial Service Providers: Focus on Real World Data Based Trial Services, Virtual Trials, Adaptive, Umbrella and Basket Design, 2019-2050. Roots Analysis. 2019 Apr. URL: https://tinyurl.com/hhse4u2d [accessed 2021-06-22]
49. Latulippe K, Hamel C, Giroux D. Social Health Inequalities and eHealth: A Literature Review With Qualitative Synthesis of Theoretical and Empirical Studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
50. Levitan B, Getz K, Eisenstein EL, Goldberg M, Harker M, Hesterlee S, et al. Assessing the Financial Value of Patient Engagement: A Quantitative Approach from CTTI's Patient Groups and Clinical Trials Project. Ther Innov Regul Sci 2018 Mar 30;52(2):220-229 [FREE Full text] [doi: 10.1177/2168479017716715] [Medline: 29714515]
51. Hirsch IB, Martinez J, Dorsey ER, Finken G, Fleming A, Gropp C, et al. Incorporating Site-less Clinical Trials Into Drug Development: A Framework for Action. Clin Ther 2017 May;39(5):1064-1076 [FREE Full text] [doi: 10.1016/j.clinthera.2017.03.018] [Medline: 28413148]
52. Multicriteria Analysis: A Manual. Wetherby, West Yorkshire, UK: Communities and Local Government Publications; 2009.
53. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012 Aug 06;13:132 [FREE Full text] [doi: 10.1186/1745-6215-13-132] [Medline: 22867278]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research

XSL•FO
RenderX

CRO: clinical research organization
CTTI: Clinical Trials Transformation Initiative
FDA: Food and Drug Administration
MCDA: multicriteria decision analysis
TNO: Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek
VCT: virtual clinical trial

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by B Johnson, E Izmailova, C Fitzer-Attas; comments to author 08.02.21; revised version received 31.03.21; accepted 06.05.21; published 06.07.21.

Please cite as:
Coert RMH, Timmis JK, Boorsma A, Pasman WJ
Stakeholder Perspectives on Barriers and Facilitators for the Adoption of Virtual Clinical Trials: Qualitative Study
J Med Internet Res 2021;23(7):e26813
URL: https://www.jmir.org/2021/7/e26813
doi:10.2196/26813
PMID:34255673

©Romée Melanie Helena Coert, James Kenneth Timmis, André Boorsma, Wilrike J Pasman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX